

10/ 077 130  
07-30-04

BC

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
25 April 2002 (25.04.2002)

PCT

(10) International Publication Number  
WO 02/33099 A2

(51) International Patent Classification<sup>7</sup>: C12N 15/54,  
9/12, A01K 67/027, C07K 16/40, C12Q 1/68, 1/48, G01N  
33/50, A61K 38/45

(21) International Application Number: PCT/US01/47728

(22) International Filing Date: 20 October 2001 (20.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/242,410 20 October 2000 (20.10.2000) US  
60/244,068 27 October 2000 (27.10.2000) US  
60/245,708 3 November 2000 (03.11.2000) US  
60/247,672 9 November 2000 (09.11.2000) US  
60/249,565 16 November 2000 (16.11.2000) US  
60/252,730 22 November 2000 (22.11.2000) US  
60/250,807 1 December 2000 (01.12.2000) US

(71) Applicant (for all designated States except US): INCYTE  
GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo  
Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): GURURAJAN,  
Rajagopal [US/US]; 5591 Dent Avenue, San Jose, CA  
95118 (US). BAUGHN, Mariah, R. [US/US]; 14244  
Santiago Road, San Leandro, CA 94577 (US). WALIA,  
Narinder, K. [US/US]; 890 David Street, #205, San Leand-  
ro, CA 94577 (US). ELLIOTT, Vicki, S. [US/US]; 3770  
Polton Place Way, San Jose, CA 95121 (US). XU, Yuming  
[US/US]; 1739 Walnut Drive, Mountain View, CA 94040  
(US). ARVIZU, Chandra [US/US]; 490 Sherwood Way  
#1, Menlo Park, CA 94025 (US). YAO, Monique, G.  
[US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US).  
RAMKUMAR, Jaya [IN/US]; 34359 Maybird Circle,  
Fremont, CA 94555 (US). DING, Li [CN/US]; 3353 Alma  
Street #146, Palo Alto, CA 94306 (US). TANG, Y., Tom  
[US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US).  
HAFALIA, April, J., A. [US/US]; 2227 Calle de Primavera,  
Santa Clara, CA 95054 (US). NGUYEN, Danniell,  
B. [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118  
(US). GANDHI, Ameena, R. [US/US]; 705 5th Avenue,  
San Francisco, CA 94118 (US). LU, Yan [CN/US];  
3885 Cornia Way, Palo Alto, CA 94303 (US). YUE,

Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087  
(US). BURFORD, Neil [GB/US]; 105 Wildwood Circle,  
Durham, CT 06422 (US). BANDMAN, Olga [US/US];  
366 Anna Avenue, Mountain View, CA 94043 (US). TRI-  
BOULEY, Catherine, M. [FR/US]; 1121 Tennessee, #5,  
San Francisco, CA 94107 (US). LAL, Preeti, G. [IN/US];  
P.O. Box 5142, Santa Clara, CA 95056 (US). RECIPON,  
Shirley, A. [US/US]; 85 Fortuna Avenue, San Francisco,  
CA 94115 (US). LU, Dyung, Aina, M. [US/US]; 233 Coy  
Drive, San Jose, CA 95123 (US). BOROWSKY, Mark,  
L. [US/US]; 122 Orchard Avenue, Redwood City, CA  
94061 (US). THORNTON, Michael [US/US]; 9 Medway  
Road, Woodside, CA 94062 (US). SWARNAKER, Anita  
[CA/US]; 8 Locksley Avenue, #5D, San Francisco, CA  
94122 (US). THANGAVELU, Kavitha [IN/US]; 1950  
Montecito Avenue #23, Mountain View, CA 94304 (US).  
KHAN, Farrah, A. [IN/US]; 3617 Central Road, #102,  
Glenview, IL 60025 (US). ISON, Craig, H. [US/US];  
1242 Weathersfield Way, San Jose, CA 95118 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics,  
Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,  
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,  
ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,  
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,  
TG).

Published:

— without international search report and to be republished  
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: HUMAN KINASES

(57) Abstract: The invention provides human human kinases (PKIN) and polynucleotides which identify and encode PKIN. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of PKIN.



WO 02/33099 A2

## HUMAN KINASES

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human kinases and to the  
5 use of these sequences in the diagnosis, treatment, and prevention of cancer, immune disorders,  
disorders affecting growth and development, cardiovascular diseases, and lipid disorders, and in the  
assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid  
sequences of human kinases.

### 10 BACKGROUND OF THE INVENTION

Kinases comprise the largest known enzyme superfamily and vary widely in their target  
molecules. Kinases catalyze the transfer of high energy phosphate groups from a phosphate donor to  
a phosphate acceptor. Nucleotides usually serve as the phosphate donor in these reactions, with most  
kinases utilizing adenosine triphosphate (ATP). The phosphate acceptor can be any of a variety of  
15 molecules, including nucleosides, nucleotides, lipids, carbohydrates, and proteins. Proteins are  
phosphorylated on hydroxyamino acids. Addition of a phosphate group alters the local charge on the  
acceptor molecule, causing internal conformational changes and potentially influencing intermolecular  
contacts. Reversible protein phosphorylation is the primary method for regulating protein activity in  
eukaryotic cells. In general, proteins are activated by phosphorylation in response to extracellular  
20 signals such as hormones, neurotransmitters, and growth and differentiation factors. The activated  
proteins initiate the cell's intracellular response by way of intracellular signaling pathways and second  
messenger molecules such as cyclic nucleotides, calcium-calmodulin, inositol, and various mitogens,  
that regulate protein phosphorylation.

Kinases are involved in all aspects of a cell's function, from basic metabolic processes, such  
25 as glycolysis, to cell-cycle regulation, differentiation, and communication with the extracellular  
environment through signal transduction cascades. Inappropriate phosphorylation of proteins in cells  
has been linked to changes in cell cycle progression and cell differentiation. Changes in the cell cycle  
have been linked to induction of apoptosis or cancer. Changes in cell differentiation have been linked  
to diseases and disorders of the reproductive system, immune system, and skeletal muscle.

30 There are two classes of protein kinases. One class, protein tyrosine kinases (PTKs),  
phosphorylates tyrosine residues, and the other class, protein serine/threonine kinases (STKs),  
phosphorylates serine and threonine residues. Some PTKs and STKs possess structural  
characteristics of both families and have dual specificity for both tyrosine and serine/threonine

residues. Almost all kinases contain a conserved 250-300 amino acid catalytic domain containing specific residues and sequence motifs characteristic of the kinase family. The protein kinase catalytic domain can be further divided into 11 subdomains. N-terminal subdomains I-IV fold into a two-lobed structure which binds and orients the ATP donor molecule, and subdomain V spans the two lobes. C-terminal subdomains VI-XI bind the protein substrate and transfer the gamma phosphate from ATP to the hydroxyl group of a tyrosine, serine, or threonine residue. Each of the 11 subdomains contains specific catalytic residues or amino acid motifs characteristic of that subdomain. For example, subdomain I contains an 8-amino acid glycine-rich ATP binding consensus motif, subdomain II contains a critical lysine residue required for maximal catalytic activity, and subdomains VI through IX comprise the highly conserved catalytic core. PTKs and STKs also contain distinct sequence motifs in subdomains VI and VIII which may confer hydroxyamino acid specificity.

In addition, kinases may also be classified by additional amino acid sequences, generally between 5 and 100 residues, which either flank or occur within the kinase domain. These additional amino acid sequences regulate kinase activity and determine substrate specificity. (Reviewed in Hardie, G. and Hanks, S. (1995) The Protein Kinase Facts Book, Vol I p.p. 17-20 Academic Press, San Diego, CA.). In particular, two protein kinase signature sequences have been identified in the kinase domain, the first containing an active site lysine residue involved in ATP binding, and the second containing an aspartate residue important for catalytic activity. If a protein analyzed includes the two protein kinase signatures, the probability of that protein being a protein kinase is close to 100% (PROSITE: PDOC00100, November 1995).

#### Protein Tyrosine Kinases

Protein tyrosine kinases (PTKs) may be classified as either transmembrane, receptor PTKs or nontransmembrane, nonreceptor PTK proteins. Transmembrane tyrosine kinases function as receptors for most growth factors. Growth factors bind to the receptor tyrosine kinase (RTK), which causes the receptor to phosphorylate itself (autophosphorylation) and specific intracellular second messenger proteins. Growth factors (GF) that associate with receptor PTKs include epidermal GF, platelet-derived GF, fibroblast GF, hepatocyte GF, insulin and insulin-like GFs, nerve GF, vascular endothelial GF, and macrophage colony stimulating factor.

Nontransmembrane, nonreceptor PTKs lack transmembrane regions and, instead, form signaling complexes with the cytosolic domains of plasma membrane receptors. Receptors that function through non-receptor PTKs include those for cytokines and hormones (growth hormone and prolactin), and antigen-specific receptors on T and B lymphocytes.

Many PTKs were first identified as oncogene products in cancer cells in which PTK

activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode PTKs. Furthermore, cellular transformation (oncogenesis) is often accompanied by increased tyrosine phosphorylation activity (Charbonneau, H. and Tonks, N. K. (1992) *Annu. Rev. Cell Biol.* 8:463-93). Regulation of PTK activity may therefore be an important strategy in controlling some types of cancer.

#### Protein Serine/Threonine Kinases

Protein serine/threonine kinases (STKs) are nontransmembrane proteins. A subclass of STKs are known as ERKs (extracellular signal regulated kinases) or MAPs (mitogen-activated protein kinases) and are activated after cell stimulation by a variety of hormones and growth factors. Cell stimulation induces a signaling cascade leading to phosphorylation of MEK (MAP/ERK kinase) which, in turn, activates ERK via serine and threonine phosphorylation. A varied number of proteins represent the downstream effectors for the active ERK and implicate it in the control of cell proliferation and differentiation, as well as regulation of the cytoskeleton. Activation of ERK is normally transient, and cells possess dual specificity phosphatases that are responsible for its down-regulation. Also, numerous studies have shown that elevated ERK activity is associated with some cancers. Other STKs include the second messenger dependent protein kinases such as the cyclic-AMP dependent protein kinases (PKA), calcium-calmodulin (CaM) dependent protein kinases, and the mitogen-activated protein kinases (MAP); the cyclin-dependent protein kinases; checkpoint and cell cycle kinases; proliferation-related kinases; 5'-AMP-activated protein kinases; and kinases involved in apoptosis.

The second messenger dependent protein kinases primarily mediate the effects of second messengers such as cyclic AMP (cAMP), cyclic GMP, inositol triphosphate, phosphatidylinositol, 3,4,5-triphosphate, cyclic ADPribose, arachidonic acid, diacylglycerol and calcium-calmodulin. The PKAs are involved in mediating hormone-induced cellular responses and are activated by cAMP produced within the cell in response to hormone stimulation. cAMP is an intracellular mediator of hormone action in all animal cells that have been studied. Hormone-induced cellular responses include thyroid hormone secretion, cortisol secretion, progesterone secretion, glycogen breakdown, bone resorption, and regulation of heart rate and force of heart muscle contraction. PKA is found in all animal cells and is thought to account for the effects of cAMP in most of these cells. Altered PKA expression is implicated in a variety of disorders and diseases including cancer, thyroid disorders, diabetes, atherosclerosis, and cardiovascular disease (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, New York, NY, pp. 416-431, 1887).

The casein kinase I (CKI) gene family is another subfamily of serine/threonine protein

kinases. This continuously expanding group of kinases have been implicated in the regulation of numerous cytoplasmic and nuclear processes, including cell metabolism, and DNA replication and repair. CKI enzymes are present in the membranes, nucleus, cytoplasm and cytoskeleton of eukaryotic cells, and on the mitotic spindles of mammalian cells (Fish, K.J. et al., (1995) J. Biol. Chem. 270:14875-14883.

The CKI family members all have a short amino-terminal domain of 9-76 amino acids, a highly conserved kinase domain of 284 amino acids, and a variable carboxyl-terminal domain that ranges from 24 to over 200 amino acids in length (Cegielska, A. et al., (1998) J. Biol. Chem. 273:1357-1364.) The CKI family is comprised of highly related proteins, as seen by the identification of isoforms of casein kinase I from a variety of sources. There are at least five mammalian isoforms,  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ , and  $\epsilon$ . Fish et al., identified CKI-epsilon from a human placenta cDNA library. It is a basic protein of 416 amino acids and is closest to CKI-delta. Through recombinant expression, it was determined to phosphorylate known CKI substrates and was inhibited by the CKI-specific inhibitor CKI-7. The human gene for CKI-epsilon was able to rescue yeast with a slow-growth phenotype caused by deletion of the yeast CKI locus, HRR250 (Fish et al, *supra*.)

The mammalian circadian mutation tau was found to be a semidominant autosomal allele of CKI-epsilon that markedly shortens period length of circadian rhythms in Syrian hamsters. The tau locus is encoded by casein kinase I-epsilon, which is also a homolog of the *Drosophila* circadian gene double-time. Studies of both the wildtype and tau mutant CKI-epsilon enzyme indicated that the mutant enzyme has a noticeable reduction in the maximum velocity and autophosphorylation state. Further, *in vitro*, CKI-epsilon is able to interact with mammalian PERIOD proteins, while the mutant enzyme is deficient in its ability to phosphorylate PERIOD. Lowrey et al., have proposed that CKI-epsilon plays a major role in delaying the negative feedback signal within the transcription-translation-based autoregulatory loop that composes the core of the circadian mechanism. Therefore the CKI-epsilon enzyme is an ideal target for pharmaceutical compounds influencing circadian rhythms, jet-lag and sleep, in addition to other physiologic and metabolic processes under circadian regulation (Lowrey, P.L. et al., (2000) Science 288:483-491.)

#### Calcium-Calmodulin Dependent Protein Kinases

Calcium-calmodulin dependent (CaM) kinases are involved in regulation of smooth muscle contraction, glycogen breakdown (phosphorylase kinase), and neurotransmission (CaM kinase I and CaM kinase II). CaM dependent protein kinases are activated by calmodulin, an intracellular calcium receptor, in response to the concentration of free calcium in the cell. Many CaM kinases are also activated by phosphorylation. Some CaM kinases are also activated by autophosphorylation or by

other regulatory kinases. CaM kinase I phosphorylates a variety of substrates including the neurotransmitter-related proteins synapsin I and II, the gene transcription regulator, CREB, and the cystic fibrosis conductance regulator protein, CFTR (Haribabu, B. et al. (1995) EMBO Journal 14:3679-3686). CaM kinase II also phosphorylates synapsin at different sites and controls the  
5 synthesis of catecholamines in the brain through phosphorylation and activation of tyrosine hydroxylase. CaM kinase II controls the synthesis of catecholamines and serotonin, through phosphorylation/activation of tyrosine hydroxylase and tryptophan hydroxylase, respectively (Fujisawa, H. (1990) BioEssays 12:27-29). The mRNA encoding a calmodulin-binding protein kinase-like protein was found to be enriched in mammalian forebrain. This protein is associated with vesicles in both  
10 axons and dendrites and accumulates largely postnatally. The amino acid sequence of this protein is similar to CaM-dependent STKs, and the protein binds calmodulin in the presence of calcium (Godbout, M. et al. (1994) J. Neurosci. 14:1-13).

#### Mitogen-Activated Protein Kinases

The mitogen-activated protein kinases (MAP) which mediate signal transduction from the cell  
15 surface to the nucleus via phosphorylation cascades are another STK family that regulates intracellular signaling pathways. Several subgroups have been identified, and each manifests different substrate specificities and responds to distinct extracellular stimuli (Egan, S.E. and Weinberg, R.A. (1993) Nature 365:781-783). MAP kinase signaling pathways are present in mammalian cells as well as in yeast. The extracellular stimuli which activate MAP kinase pathways include epidermal growth  
20 factor (EGF), ultraviolet light, hyperosmolar medium, heat shock, endotoxic lipopolysaccharide (LPS), and pro-inflammatory cytokines such as tumor necrosis factor (TNF) and interleukin-1 (IL-1). Altered MAP kinase expression is implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development.

#### Cyclin-Dependent Protein Kinases

25 The cyclin-dependent protein kinases (CDKs) are STKs that control the progression of cells through the cell cycle. The entry and exit of a cell from mitosis are regulated by the synthesis and destruction of a family of activating proteins called cyclins. Cyclins are small regulatory proteins that bind to and activate CDKs, which then phosphorylate and activate selected proteins involved in the mitotic process. CDKs are unique in that they require multiple inputs to become activated. In addition  
30 to cyclin binding, CDK activation requires the phosphorylation of a specific threonine residue and the dephosphorylation of a specific tyrosine residue on the CDK.

Another family of STKs associated with the cell cycle are the NIMA (never in mitosis)-related kinases (Neks). Both CDKs and Neks are involved in duplication, maturation, and separation

of the microtubule organizing center, the centrosome, in animal cells (Fry, A.M., et al. (1998) EMBO J. 17:470-481). The NIM-related kinases also include NIK1 histidine kinases, which function in signal transmission (Yamada-Okabe, T. et al. (1999) J. Bacteriol. 181:7243-7247).

#### Checkpoint and Cell Cycle Kinases

5 In the process of cell division, the order and timing of cell cycle transitions are under control of cell cycle checkpoints, which ensure that critical events such as DNA replication and chromosome segregation are carried out with precision. If DNA is damaged, e.g. by radiation, a checkpoint pathway is activated that arrests the cell cycle to provide time for repair. If the damage is extensive, apoptosis is induced. In the absence of such checkpoints, the damaged DNA is inherited by aberrant  
10 cells which may cause proliferative disorders such as cancer. Protein kinases play an important role in this process. For example, a specific kinase, checkpoint kinase 1 (Chk1), has been identified in yeast and mammals, and is activated by DNA damage in yeast. Activation of Chk1 leads to the arrest of the cell at the G2/M transition. (Sanchez, Y. et al. (1997) Science 277:1497-1501.) Specifically, Chk1 phosphorylates the cell division cycle phosphatase CDC25, inhibiting its normal function which is  
15 to dephosphorylate and activate the cyclin-dependent kinase Cdc2. Cdc2 activation controls the entry of cells into mitosis. (Peng, C-Y et al. (1997) Science 277:1501- 1505.) Thus, activation of Chk1 prevents the damaged cell from entering mitosis. A similar deficiency in a checkpoint kinase, such as Chk1, may also contribute to cancer by failure to arrest cells with damaged DNA at other checkpoints such as G2/M.

#### 20 Proliferation-Related Kinases

Proliferation-related kinase is a serum/cytokine inducible STK that is involved in regulation of the cell cycle and cell proliferation in human megakaryocytic cells (Li, B. et al. (1996) J. Biol. Chem. 271:19402-8). Proliferation-related kinase is related to the polo (derived from *Drosophila* polo gene) family of STKs implicated in cell division. Proliferation-related kinase is downregulated in lung tumor  
25 tissue and may be a proto-oncogene whose deregulated expression in normal tissue leads to oncogenic transformation.

The RET (rearranged during transfection) proto-oncogene encodes a tyrosine kinase receptor involved in both multiple endocrine neoplasia type 2, an inherited cancer syndrome, and Hirschsprung disease, a developmental defect of enteric neurons. RET and its functional ligand, glial cell line-  
30 derived neurotrophic factor, play key roles in the development of the human enteric nervous system (Pachnis, V. et al. (1998) Am. J. Physiol. 275:G183-G186).

#### 5'-AMP-activated protein kinase

A ligand-activated STK protein kinase is 5'-AMP-activated protein kinase (AMPK) (Gao, G.

et al. (1996) J. Biol Chem. 271:8675-8681). Mammalian AMPK is a regulator of fatty acid and sterol synthesis through phosphorylation of the enzymes acetyl-CoA carboxylase and hydroxymethylglutaryl-CoA reductase and mediates responses of these pathways to cellular stresses such as heat shock and depletion of glucose and ATP. AMPK is a heterotrimeric complex comprised of a catalytic alpha subunit and two non-catalytic beta and gamma subunits that are believed to regulate the activity of the alpha subunit. Subunits of AMPK have a much wider distribution in non-lipogenic tissues such as brain, heart, spleen, and lung than expected. This distribution suggests that its role may extend beyond regulation of lipid metabolism alone.

#### Kinases in Apoptosis

Apoptosis is a highly regulated signaling pathway leading to cell death that plays a crucial role in tissue development and homeostasis. Deregulation of this process is associated with the pathogenesis of a number of diseases including autoimmune disease, neurodegenerative disorders, and cancer. Various STKs play key roles in this process. ZIP kinase is an STK containing a C-terminal leucine zipper domain in addition to its N-terminal protein kinase domain. This C-terminal domain appears to mediate homodimerization and activation of the kinase as well as interactions with transcription factors such as activating transcription factor, ATF4, a member of the cyclic-AMP responsive element binding protein (ATF/CREB) family of transcriptional factors (Sanjo, H. et al. (1998) J. Biol. Chem, 273:29066-29071). DRAK1 and DRAK2 are STKs that share homology with the death-associated protein kinases (DAP kinases), known to function in interferon- $\gamma$  induced apoptosis (Sanjo et al. supra). Like ZIP kinase, DAP kinases contain a C-terminal protein-protein interaction domain, in the form of ankyrin repeats, in addition to the N-terminal kinase domain. ZIP, DAP, and DRAK kinases induce morphological changes associated with apoptosis when transfected into NIH3T3 cells (Sanjo et al. supra). However, deletion of either the N-terminal kinase catalytic domain or the C-terminal domain of these proteins abolishes apoptosis activity, indicating that in addition to the kinase activity, activity in the C-terminal domain is also necessary for apoptosis, possibly as an interacting domain with a regulator or a specific substrate.

RICK is another STK recently identified as mediating a specific apoptotic pathway involving the death receptor, CD95 (Inohara, N. et al. (1998) J. Biol. Chem. 273:12296-12300). CD95 is a member of the tumor necrosis factor receptor superfamily and plays a critical role in the regulation and homeostasis of the immune system (Nagata, S. (1997) Cell 88:355-365). The CD95 receptor signaling pathway involves recruitment of various intracellular molecules to a receptor complex following ligand binding. This process includes recruitment of the cysteine protease caspase-8 which, in turn, activates a caspase cascade leading to cell death. RICK is composed of an N-terminal kinase



catalytic domain and a C-terminal "caspase-recruitment" domain that interacts with caspase-like domains, indicating that RICK plays a role in the recruitment of caspase-8. This interpretation is supported by the fact that the expression of RICK in human 293T cells promotes activation of caspase-8 and potentiates the induction of apoptosis by various proteins involved in the CD95 apoptosis pathway (Inohara et al. supra).

#### Mitochondrial Protein Kinases

A novel class of eukaryotic kinases, related by sequence to prokaryotic histidine protein kinases, are the mitochondrial protein kinases (MPKs) which seem to have no sequence similarity with other eukaryotic protein kinases. These protein kinases are located exclusively in the mitochondrial matrix space and may have evolved from genes originally present in respiration-dependent bacteria which were endocytosed by primitive eukaryotic cells. MPKs are responsible for phosphorylation and inactivation of the branched-chain alpha-ketoacid dehydrogenase and pyruvate dehydrogenase complexes (Harris, R.A. et al. (1995) Adv. Enzyme Regul. 34:147-162). Five MPKs have been identified. Four members correspond to pyruvate dehydrogenase kinase isozymes, regulating the activity of the pyruvate dehydrogenase complex, which is an important regulatory enzyme at the interface between glycolysis and the citric acid cycle. The fifth member corresponds to a branched-chain alpha-ketoacid dehydrogenase kinase, important in the regulation of the pathway for the disposal of branched-chain amino acids. (Harris, R.A. et al. (1997) Adv. Enzyme Regul. 37:271-293). Both starvation and the diabetic state are known to result in a great increase in the activity of the pyruvate dehydrogenase kinase in the liver, heart and muscle of the rat. This increase contributes in both disease states to the phosphorylation and inactivation of the pyruvate dehydrogenase complex and conservation of pyruvate and lactate for gluconeogenesis (Harris (1995) supra).

#### KINASES WITH NON-PROTEIN SUBSTRATES

##### Lipid and Inositol kinases

Lipid kinases phosphorylate hydroxyl residues on lipid head groups. A family of kinases involved in phosphorylation of phosphatidylinositol (PI) has been described, each member phosphorylating a specific carbon on the inositol ring (Leevers, S.J. et al. (1999) Curr. Opin. Cell. Biol. 11:219-225). The phosphorylation of phosphatidylinositol is involved in activation of the protein kinase C signaling pathway. The inositol phospholipids (phosphoinositides) intracellular signaling pathway begins with binding of a signaling molecule to a G-protein linked receptor in the plasma membrane. This leads to the phosphorylation of phosphatidylinositol (PI) residues on the inner side of the plasma

membrane by inositol kinases, thus converting PI residues to the biphosphate state (PIP<sub>2</sub>). PIP<sub>2</sub> is then cleaved into inositol triphosphate (IP<sub>3</sub>) and diacylglycerol. These two products act as mediators for separate signaling pathways. Cellular responses that are mediated by these pathways are glycogen breakdown in the liver in response to vasopressin, smooth muscle contraction in response to acetylcholine, and thrombin-induced platelet aggregation.

PI 3-kinase (PI3K), which phosphorylates the D3 position of PI and its derivatives, has a central role in growth factor signal cascades involved in cell growth, differentiation, and metabolism. PI3K is a heterodimer consisting of an adapter subunit and a catalytic subunit. The adapter subunit acts as a scaffolding protein, interacting with specific tyrosine-phosphorylated proteins, lipid moieties, and other cytosolic factors. When the adapter subunit binds tyrosine phosphorylated targets, such as the insulin responsive substrate (IRS)-1, the catalytic subunit is activated and converts PI (4,5) biphosphate (PIP<sub>2</sub>) to PI (3,4,5) P<sub>3</sub> (PIP<sub>3</sub>). PIP<sub>3</sub> then activates a number of other proteins, including PKA, protein kinase B (PKB), protein kinase C (PKC), glycogen synthase kinase (GSK)-3, and p70 ribosomal s6 kinase. PI3K also interacts directly with the cytoskeletal organizing proteins, Rac, rho, and cdc42 (Shepherd, P.R., et al. (1998) Biochem. J. 333:471-490). Animal models for diabetes, such as *obese* and *fat* mice, have altered PI3K adapter subunit levels. Specific mutations in the adapter subunit have also been found in an insulin-resistant Danish population, suggesting a role for PI3K in type-2 diabetes (Shepard, *supra*).

PKC is also activated by diacylglycerol (DAG). Phorbol esters (PE) are analogs of DAG and tumor promoters that cause a variety of physiological changes when administered to cells and tissues. PE and DAG bind to the N-terminal region of PKC. This region contains one or more copies of a cysteine-rich domain about 50 amino-acid residues long and essential for DAG/PE-binding. Diacylglycerol kinase (DGK), the enzyme that converts DAG into phosphatidate, contains two copies of the DAG/PE-binding domain in its N-terminal section (Azzi, A. et al. (1992) Eur. J. Biochem. 208:547-557).

An example of lipid kinase phosphorylation activity is the phosphorylation of D-erythro-sphingosine to the sphingolipid metabolite, sphingosine-1-phosphate (SPP). SPP has emerged as a novel lipid second-messenger with both extracellular and intracellular actions (Kohama, T. et al. (1998) J. Biol. Chem. 273:23722-23728). Extracellularly, SPP is a ligand for the G-protein coupled receptor EDG-1 (endothelial-derived, G-protein coupled receptor). Intracellularly, SPP regulates cell growth, survival, motility, and cytoskeletal changes. SPP levels are regulated by sphingosine kinases that specifically phosphorylate D-erythro-sphingosine to SPP. The importance of sphingosine kinase in cell signaling is indicated by the fact that various stimuli, including

platelet-derived growth factor (PDGF), nerve growth factor, and activation of protein kinase C, increase cellular levels of SPP by activation of sphingosine kinase, and the fact that competitive inhibitors of the enzyme selectively inhibit cell proliferation induced by PDGF (Kohama et al. supra).

#### Purine Nucleotide Kinases

5           The purine nucleotide kinases, adenylate kinase (ATP:AMP phosphotransferase, or AdK) and guanylate kinase (ATP:GMP phosphotransferase, or GuK) play a key role in nucleotide metabolism and are crucial to the synthesis and regulation of cellular levels of ATP and GTP, respectively. These two molecules are precursors in DNA and RNA synthesis in growing cells and provide the primary source of biochemical energy in cells (ATP), and signal transduction pathways (GTP). Inhibition of  
10 various steps in the synthesis of these two molecules has been the basis of many antiproliferative drugs for cancer and antiviral therapy (Pillwein, K. et al. (1990) Cancer Res. 50:1576-1579).

AdK is found in almost all cell types and is especially abundant in cells having high rates of ATP synthesis and utilization such as skeletal muscle. In these cells AdK is physically associated with mitochondria and myofibrils, the subcellular structures that are involved in energy production and  
15 utilization, respectively. Recent studies have demonstrated a major function for AdK in transferring high energy phosphoryls from metabolic processes generating ATP to cellular components consuming ATP (Zelevnikar, R.J. et al. (1995) J. Biol. Chem. 270:7311-7319). Thus AdK may have a pivotal role in maintaining energy production in cells, particularly those having a high rate of growth or metabolism such as cancer cells, and may provide a target for suppression of its activity to treat  
20 certain cancers. Alternatively, reduced AdK activity may be a source of various metabolic, muscle-energy disorders that can result in cardiac or respiratory failure and may be treatable by increasing AdK activity.

GuK, in addition to providing a key step in the synthesis of GTP for RNA and DNA synthesis, also fulfills an essential function in signal transduction pathways of cells through the regulation of GDP  
25 and GTP. Specifically, GTP binding to membrane associated G proteins mediates the activation of cell receptors, subsequent intracellular activation of adenyl cyclase, and production of the second messenger, cyclic AMP. GDP binding to G proteins inhibits these processes. GDP and GTP levels also control the activity of certain oncogenic proteins such as p21<sup>ras</sup> known to be involved in control of cell proliferation and oncogenesis (Bos, J.L. (1989) Cancer Res. 49:4682-4689). High ratios of  
30 GTP:GDP caused by suppression of GuK cause activation of p21<sup>ras</sup> and promote oncogenesis. Increasing GuK activity to increase levels of GDP and reduce the GTP:GDP ratio may provide a therapeutic strategy to reverse oncogenesis.

GuK is an important enzyme in the phosphorylation and activation of certain antiviral drugs

useful in the treatment of herpes virus infections. These drugs include the guanine homologs acyclovir and buciclovir (Miller, W.H. and Miller R.L. (1980) J. Biol. Chem. 255:7204-7207; Stenberg, K. et al. (1986) J. Biol. Chem. 261:2134-2139). Increasing GuK activity in infected cells may provide a therapeutic strategy for augmenting the effectiveness of these drugs and possibly for reducing the necessary dosages of the drugs.

#### Pyrimidine Kinases

The pyrimidine kinases are deoxycytidine kinase and thymidine kinase 1 and 2. Deoxycytidine kinase is located in the nucleus, and thymidine kinase 1 and 2 are found in the cytosol (Johansson, M. et al. (1997) Proc. Natl. Acad. Sci. U.S.A. 94:11941-11945). Phosphorylation of deoxyribonucleosides by pyrimidine kinases provides an alternative pathway for de novo synthesis of DNA precursors. The role of pyrimidine kinases, like purine kinases, in phosphorylation is critical to the activation of several chemotherapeutically important nucleoside analogues (Arner E.S. and Eriksson, S. (1995) Pharmacol. Ther. 67:155-186).

The discovery of new human kinases, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of human kinases.

## **SUMMARY OF THE INVENTION**

The invention features purified polypeptides, human kinases, referred to collectively as "PKIN" and individually as "PKIN-1," "PKIN-2," "PKIN-3," "PKIN-4," "PKIN-5," "PKIN-6," "PKIN-7," "PKIN-8," "PKIN-9," "PKIN-10," "PKIN-11," "PKIN-12," "PKIN-13," "PKIN-14," "PKIN-15," "PKIN-16," "PKIN-17," "PKIN-18," "PKIN-19," "PKIN-20," "PKIN-21," and "PKIN-22." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-22.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from

the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-22. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:23-44.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an

amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.

The invention further provides an isolated polynucleotide selected from the group consisting of

5 a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide

10 comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least

15 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to

20 said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said

25 target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the

30 polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the

invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional PKIN, comprising administering to a patient in need of such treatment the composition.

5           The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide  
10       having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

15           The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of a polypeptide  
20       having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the  
25       polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

          The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a  
30       polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

          The invention further provides a method for assessing toxicity of a test compound, said



method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, ii) a

5 polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological

10 sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of

15 ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of

20 toxicity of the test compound.

### BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

25 Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for

30 analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

## DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## DEFINITIONS

"PKIN" refers to the amino acid sequences of substantially purified PKIN obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of PKIN. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PKIN either by directly interacting with PKIN or by acting on components of the biological pathway in which PKIN participates.

An "allelic variant" is an alternative form of the gene encoding PKIN. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to  
5 allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding PKIN include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as PKIN or a  
10 polypeptide with at least one functional characteristic of PKIN. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PKIN, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding PKIN. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino  
15 acid residues which produce a silent change and result in a functionally equivalent PKIN. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PKIN is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and  
20 arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide,  
25 polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence.  
30 Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of PKIN. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small

molecules, or any other compound or composition which modulates the activity of PKIN either by directly interacting with PKIN or by acting on components of the biological pathway in which PKIN participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind PKIN polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH<sub>2</sub>), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-

handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic PKIN, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding PKIN or fragments of PKIN may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been

assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

5       “Conservative amino acid substitutions” are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
10	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
15	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
	His	Asn, Arg, Gln, Glu
20	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
25	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
30	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

35       A “deletion” refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term “derivative” refers to a chemically modified polynucleotide or polypeptide.

Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains  
40 at least one biological or immunological function of the natural molecule. A derivative polypeptide is

one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A “detectable label” refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

5 “Differential expression” refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

10 “Exon shuffling” refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A “fragment” is a unique portion of PKIN or the polynucleotide encoding PKIN which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

25 A fragment of SEQ ID NO:23-44 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:23-44, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:23-44 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:23-44 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:23-44 and the region of SEQ ID NO:23-44 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

30 A fragment of SEQ ID NO:1-22 is encoded by a fragment of SEQ ID NO:23-44. A fragment of SEQ ID NO:1-22 comprises a region of unique amino acid sequence that specifically

identifies SEQ ID NO:1-22. For example, a fragment of SEQ ID NO:1-22 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-22. The precise length of a fragment of SEQ ID NO:1-22 and the region of SEQ ID NO:1-22 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the  
5 intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two  
10 or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore  
15 achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in  
20 Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

25 Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis  
30 programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The



"BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

5        *Matrix: BLOSUM62*  
        *Reward for match: 1*  
        *Penalty for mismatch: -2*  
        *Open Gap: 5 and Extension Gap: 2 penalties*  
        *Gap x drop-off: 50*  
 10       *Expect: 10*  
        *Word Size: 11*  
        *Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example,  
 15 over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

20 Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to  
 25 the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

30 Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap

penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise  
5 comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Open Gap: 11 and Extension Gap: 1 penalties*

10 *Gap x drop-off: 50*

*Expect: 10*

*Word Size: 3*

*Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence,  
15 for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be  
20 used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid  
25 sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity.  
30 Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive

conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 5 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of 10 the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention 15 include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as 20 formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

25 The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g.,  $C_0t$  or  $R_0t$  analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate 30 to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune

disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of PKIN which is  
5 capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of PKIN which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

10 The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of PKIN. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PKIN.

15 The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a  
20 functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which  
25 comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an PKIN may involve lipidation, glycosylation,  
30 phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of PKIN.

"Probe" refers to nucleic acid sequences encoding PKIN, their complements, or fragments

thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target  
5 polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also  
10 be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for  
15 example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that  
20 purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000  
25 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer  
30 selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may

also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned  
5 nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are  
10 not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques  
15 such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a  
20 vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription,  
25 translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

30 An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing PKIN, nucleic acids encoding PKIN, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

5       The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the  
10       epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

      The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with  
15       which they are naturally associated.

      A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

      "Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers,  
20       microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

      A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

      "Transformation" describes a process by which exogenous DNA is introduced into a recipient  
25       cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed  
30       cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

      A "transgenic organism," as used herein, is any organism, including but not limited to animals

and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at



least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

## 5 THE INVENTION

The invention is based on the discovery of new human human kinases (PKIN), the polynucleotides encoding PKIN, and the use of these compositions for the diagnosis, treatment, or prevention of cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders.

10 Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is  
15 denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte  
20 polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

25 Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS  
30 program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are human kinases.

For example, SEQ ID NO:1 is 91% identical to human casein kinase I-alpha (GenBank ID g852055) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $2.9e-167$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a eukaryotic protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:1 is a protein kinase.

For example, SEQ ID NO:10 is 91% identical to Mus musculus FYVE finger-containing phosphoinositide kinase (GenBank ID g4200446) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:10 also contains a phosphatidyl inositol 4-phosphate 5-kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from PRODOM analysis provides further corroborative evidence that SEQ ID NO:10 is a phosphoinositide kinase.

For example, SEQ ID NO:12 is 71% identical to human serine/threonine protein kinase (GenBank ID g7160989) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $1.7e-148$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:12 also contains a eukaryotic protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:12 is protein kinase.

For example, SEQ ID NO:13 is 86% identical to murine pantothenate kinase 1 beta (GenBank ID g6690020) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $1.6e-129$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. Pantothenate kinase (PanK) is proposed to be the master regulator of CoA biosynthesis in mammalian cells, by controlling flux through the CoA biosynthetic pathway. Changes in the level of tissue PanK activity is reflected by the concurrent changes in the levels of CoA as seen in various metabolic states. Alterations in CoA levels and PanK activity are

seen during starvation/feeding, pathological states such as diabetes and by treatment with hypolipidemic drugs (Rock, C.O. et al., (2000) J. Biol. Chem. 275:1377-1383.)

For example, SEQ ID NO:16 is 68% identical to Mus musculus Nck-interacting kinase-like embryo specific kinase (GenBank ID g6472874) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:16 also contains a eukaryotic protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:16 is a protein kinase.

For example, SEQ ID NO:19 is 99% identical to human RET tyrosine kinase receptor (GenBank ID g5419753) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:19 also contains a eukaryotic protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:19 is a tyrosine kinase.

For example, SEQ ID NO:22 is 33% identical to *Gallus gallus* smooth muscle myosin light chain kinase precursor (GenBank ID g212661) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.2 e-60, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:22 also contains two eukaryotic protein kinase domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:22 is a protein kinase.

SEQ ID NO:2-9, SEQ ID NO:11, SEQ ID NO:14-15, SEQ ID NO:17-18, and SEQ ID NO:20-21 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-22 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide

consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:23-44 or that distinguish between SEQ ID NO:23-44 and  
 5 related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective  
 10 full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 183812R7 is the identification number of an Incyte cDNA sequence, and CARDNOT01 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from  
 15 pooled cDNA libraries (e.g., 71583296V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 may be derived from the  
 20 NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL\_XXXXXX\_N<sub>1</sub>\_N<sub>2</sub>\_YYYY\_N<sub>3</sub>\_N<sub>4</sub> represents a "stitched" sequence in  
 25 which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYY is the number of the prediction generated by the algorithm, and N<sub>1,2,3...</sub>, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FLXXXXXX\_gAAAAA\_gBBBBB\_1\_N is the  
 30 identification number of a "stretched" sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to

specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK)
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses PKIN variants. A preferred PKIN variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the PKIN amino acid sequence, and which contains at least one functional or structural characteristic of PKIN.

The invention also encompasses polynucleotides which encode PKIN. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:23-44, which encodes PKIN. The polynucleotide sequences of SEQ ID NO:23-44, as presented in the Sequence Listing, embrace the equivalent RNA sequences,

wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding PKIN. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding PKIN. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:23-44 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:23-44. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PKIN.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PKIN, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PKIN, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode PKIN and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring PKIN under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PKIN or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PKIN and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode PKIN and PKIN derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PKIN or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:23-44 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding PKIN may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.)

Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060.)

Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National

- 5 Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include  
10 sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary  
15 sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer  
20 controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode PKIN may be cloned in recombinant DNA molecules that direct expression of PKIN, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy  
25 of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PKIN.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter PKIN-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA  
30 shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.



The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of PKIN, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding PKIN may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, PKIN itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of PKIN, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active PKIN, the nucleotide sequences encoding PKIN or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in

a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PKIN. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PKIN. Such signals  
5 include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding PKIN and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided  
10 by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression  
15 vectors containing sequences encoding PKIN and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

20 A variety of expression vector/host systems may be utilized to contain and express sequences encoding PKIN. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or  
25 tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New  
30 York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola,

M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) The invention is not limited by the host cell employed.

5           In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding PKIN. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding PKIN can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPO1 plasmid (Life Technologies). Ligation of sequences encoding PKIN into the vector's multiple cloning  
10 site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) When large quantities of PKIN are needed, e.g. for the production of  
15 antibodies, vectors which direct high level expression of PKIN may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

          Yeast expression systems may be used for production of PKIN. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such  
20 vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, *supra*; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; and Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184.)

          Plant systems may also be used for expression of PKIN. Transcription of sequences  
25 encoding PKIN may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These  
30 constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., *The McGraw Hill Yearbook of Science and Technology* (1992) McGraw Hill, New York NY, pp. 191-196.)

          In mammalian cells, a number of viral-based expression systems may be utilized. In cases

where an adenovirus is used as an expression vector, sequences encoding PKIN may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses PKIN in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of PKIN in cell lines is preferred. For example, sequences encoding PKIN can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *ap<sup>r</sup>* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate

luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) *Methods Mol. Biol.* 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest  
5 is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding PKIN is inserted within a marker gene sequence, transformed cells containing sequences encoding PKIN can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding PKIN under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates  
10 expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding PKIN and that express PKIN may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or  
15 chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of PKIN using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing  
20 monoclonal antibodies reactive to two non-interfering epitopes on PKIN is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) *Serological Methods, a Laboratory Manual*, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) *Current Protocols in Immunology*, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) *Immunochemical Protocols*, Humana Press,  
25 Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PKIN include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide.  
30 Alternatively, the sequences encoding PKIN, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety

of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

5           Host cells transformed with nucleotide sequences encoding PKIN may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode PKIN may be designed to contain signal sequences which direct  
10   secretion of PKIN through a prokaryotic or eukaryotic cell membrane.

          In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the  
15   protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

20           In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PKIN may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric PKIN protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of PKIN activity. Heterologous protein and  
25   peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins,  
30   respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the PKIN encoding sequence and the heterologous protein sequence, so that PKIN

may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled PKIN may be achieved in  
5 vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

PKIN of the present invention or fragments thereof may be used to screen for compounds  
10 that specifically bind to PKIN. At least one and up to a plurality of test compounds may be screened for specific binding to PKIN. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of PKIN, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a  
15 natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which PKIN binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express PKIN, either as a secreted  
20 protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing PKIN or cell membrane fractions which contain PKIN are then contacted with a test compound and binding, stimulation, or inhibition of activity of either PKIN or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is  
25 detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with PKIN, either in solution or affixed to a solid support, and detecting the binding of PKIN to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural  
30 product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

PKIN of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of PKIN. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for PKIN

activity, wherein PKIN is combined with at least one test compound, and the activity of PKIN in the presence of a test compound is compared with the activity of PKIN in the absence of the test compound. A change in the activity of PKIN in the presence of the test compound is indicative of a compound that modulates the activity of PKIN. Alternatively, a test compound is combined with an in vitro or cell-free system comprising PKIN under conditions suitable for PKIN activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of PKIN may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding PKIN or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding PKIN may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding PKIN can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding PKIN is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with



potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress PKIN, e.g., by secreting PKIN in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

## THERAPEUTICS

- 5           Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of PKIN and human kinases. In addition, the expression of PKIN is closely associated with brain, breast tumor, cardiovascular, digestive, fallopian tube tumor, fetal stomach, nervous, ovarian tumor, pancreatic tumor, peritoneal tumor, pituitary gland, placental, prostate tumor, neural, spinal cord, and testicular tissues, and with umbilical cord blood dendritic cells. Therefore, PKIN appears to play a
- 10          role in cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders. In the treatment of disorders associated with increased PKIN expression or activity, it is desirable to decrease the expression or activity of PKIN. In the treatment of disorders associated with decreased PKIN expression or activity, it is desirable to increase the expression or activity of PKIN.
- 15           Therefore, in one embodiment, PKIN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia,
- 20          gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease; an immune disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune
- 25          thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis,
- 30          myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral,

bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a growth and developmental disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including

5 adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy,

10 gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and

15 sensorineural hearing loss; a cardiovascular disease, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease,

20 degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung

25 anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic

30 pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and

noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; and a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's  
5 disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM<sub>2</sub> gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia,  
10 primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity.

In another embodiment, a vector capable of expressing PKIN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased  
15 expression or activity of PKIN including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified PKIN in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN including, but not limited to, those provided above.

20 In still another embodiment, an agonist which modulates the activity of PKIN may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN including, but not limited to, those listed above.

In a further embodiment, an antagonist of PKIN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PKIN. Examples of such  
25 disorders include, but are not limited to, those cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders described above. In one aspect, an antibody which specifically binds PKIN may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PKIN.

30 In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PKIN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PKIN including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary

sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of PKIN may be produced using methods which are generally known in the art. In particular, purified PKIN may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PKIN. Antibodies to PKIN may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with PKIN or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PKIN have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of PKIN amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to PKIN may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the

splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single  
5 chain antibodies may be adapted, using methods known in the art, to produce PKIN-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population  
10 or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for PKIN may also be generated. For example, such fragments include, but are not limited to, F(ab)<sub>2</sub> fragments produced by pepsin  
15 digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab)<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired  
20 specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PKIN and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering PKIN epitopes is generally used, but a competitive binding assay may also be  
25 employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for PKIN. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of PKIN-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$   
30 determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple PKIN epitopes, represents the average affinity, or avidity, of the antibodies for PKIN. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular PKIN epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$

ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the PKIN-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_d$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of PKIN, preferably in active form, from the antibody

5 (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a

10 polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of PKIN-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

15 In another embodiment of the invention, the polynucleotides encoding PKIN, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding PKIN. Such technology is well known in the art, and antisense oligonucleotides or larger fragments

20 can be designed from various locations along the coding or control regions of sequences encoding PKIN. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence

25 complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other

30 gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding PKIN may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in PKIN expression or regulation causes disease, the expression of PKIN from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in PKIN are treated by constructing mammalian expression vectors encoding PKIN and introducing these vectors by mechanical means into PKIN-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J-L. and H. Récipon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of PKIN include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). PKIN may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter

(e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the  
5 FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding PKIN from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver  
10 polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

15 In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to PKIN expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding PKIN under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences  
20 required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al.  
25 (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference.  
30 Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998)



Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding PKIN to cells which have one or more genetic abnormalities with respect to the expression of PKIN. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding PKIN to target cells which have one or more genetic abnormalities with respect to the expression of PKIN. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing PKIN to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding PKIN to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based

on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for PKIN into the alphavirus genome in place of the capsid-coding region results in the production of a large number of PKIN-coding RNAs and the synthesis of high levels of PKIN in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of PKIN into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding PKIN.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for

secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared  
5 by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding PKIN. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs  
10 that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages  
15 within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a  
20 compound which is effective in altering expression of a polynucleotide encoding PKIN. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective  
25 compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased PKIN expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding PKIN may be therapeutically useful, and in the treatment of disorders associated with decreased PKIN expression or activity, a compound which specifically promotes expression of the  
30 polynucleotide encoding PKIN may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in

altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a  
5 polynucleotide encoding PKIN is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding PKIN are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the  
10 polynucleotide encoding PKIN. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide  
15 can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified  
20 oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient.  
25 Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and  
30 monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various

formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of PKIN, antibodies to PKIN, and mimetics, agonists, antagonists, or inhibitors of PKIN.

The compositions utilized in this invention may be administered by any number of routes  
5 including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the  
10 case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without  
15 needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of  
20 macromolecules comprising PKIN or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, PKIN or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et  
25 al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for  
30 administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example PKIN or fragments thereof, antibodies of PKIN, and agonists, antagonists or inhibitors of PKIN, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by

standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Compositions which exhibit large  
5 therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

10 The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response  
15 to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art.  
20 Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind PKIN may be used for the  
25 diagnosis of disorders characterized by expression of PKIN, or in assays to monitor patients being treated with PKIN or agonists, antagonists, or inhibitors of PKIN. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for PKIN include methods which utilize the antibody and a label to detect PKIN in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and  
30 may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring PKIN, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PKIN expression. Normal

or standard values for PKIN expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to PKIN under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of PKIN expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PKIN may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of PKIN may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of PKIN, and to monitor regulation of PKIN levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding PKIN or closely related molecules may be used to identify nucleic acid sequences which encode PKIN. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding PKIN, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the PKIN encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:23-44 or from genomic sequences including promoters, enhancers, and introns of the PKIN gene.

Means for producing specific hybridization probes for DNAs encoding PKIN include the cloning of polynucleotide sequences encoding PKIN or PKIN derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding PKIN may be used for the diagnosis of disorders associated with expression of PKIN. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma,

teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, leukemias such as multiple myeloma and lymphomas such as Hodgkin's disease; an immune disorder, 5 such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, 10 erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, 15 thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a growth and developmental disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and 20 cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker 25 muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, 30 craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a cardiovascular disease, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary



artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic

5 endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary

10 disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary

15 hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; and a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid

20 storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM<sub>2</sub> gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with

25 hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity. The polynucleotide sequences encoding PKIN may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick,

30 pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PKIN expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding PKIN may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide

sequences encoding PKIN may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding PKIN in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of PKIN, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PKIN, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding PKIN may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding PKIN, or a fragment of a polynucleotide complementary to the polynucleotide encoding PKIN, and will be employed under optimized conditions for identification of a specific gene or

condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding PKIN may be used to detect single nucleotide polymorphisms (SNPs). SNPs are  
5 substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding PKIN are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal  
10 tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP  
15 (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the  
20 high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of PKIN include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C. et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be  
25 accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray  
30 can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor

progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, PKIN, fragments of PKIN, or antibodies specific for PKIN may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

10 A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 15 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The 20 resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention 25 may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) 30 Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression

provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a  
5 toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed  
10 gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be  
15 quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global  
20 pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is  
25 achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot  
30 is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for

example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein  
5 identification.

A proteomic profile may also be generated using antibodies specific for PKIN to quantify the levels of PKIN expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lucking, A. et al. (1999) *Anal. Biochem.* 270:103-  
10 111; Mendoz, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and  
15 should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid  
20 degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample.  
25 A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological  
30 sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the

two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding PKIN may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding PKIN on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery

techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant  
5 invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, PKIN, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a  
10 solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PKIN and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are  
15 synthesized on a solid substrate. The test compounds are reacted with PKIN, or fragments thereof, and washed. Bound PKIN is then detected by methods well known in the art. Purified PKIN can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

20 In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PKIN specifically compete with a test compound for binding PKIN. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PKIN.

In additional embodiments, the nucleotide sequences which encode PKIN may be used in any  
25 molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore,  
30 to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific



embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 60/242,410, U.S. Ser. No. 60/244,068, U.S. Ser. No. 60/245,708, U.S. Ser. No. 60/247,672, U.S. Ser. No. 60/249,565, U.S. Ser. No. 60/252,730, and U.S. Ser. No. 60/250,807, are hereby expressly incorporated by reference.

## EXAMPLES

### I. Construction of cDNA Libraries

10 Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl  
15 cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles  
20 (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the  
25 UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-  
30 1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid

(Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or  
5 ElectroMAX DH10B from Life Technologies.

## II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Miniprep DNA purification system (Promega); an  
10 AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4 °C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a  
15 high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

## 20 III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the  
25 MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the  
30 ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the

techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the

identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:23-44. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and  
5 amplification technologies are described in Table 4, column 4.

#### IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative human kinases were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of  
10 organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA  
15 sequences encode human kinases, the encoded polypeptides were analyzed by querying against PFAM models for human kinases. Potential human kinases were also identified by homology to Incyte cDNA sequences that had been annotated as human kinases. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to  
20 the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted  
25 coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

#### V. Assembly of Genomic Sequence Data with cDNA Sequence Data

##### "Stitched" Sequences

30 Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm

based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpi public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

#### "Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

#### **VI. Chromosomal Mapping of PKIN Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:23-44 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:23-44 were assembled into clusters of contiguous and overlapping sequences using

assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment  
5 of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in  
10 humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease  
15 genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:29 was mapped to chromosome 1 within the interval from 199.20 to 203.00 centiMorgans, to chromosome 13 within the interval from 105.20 centiMorgans to the q terminus, and to chromosome 6 within the interval from 59.60 to 72.20 centiMorgans. More than one map location is reported for SEQ ID NO:29, indicating that sequences having different map locations  
20 were assembled into a single cluster. This situation occurs, for example, when sequences having strong similarity, but not complete identity, are assembled into a single cluster.

## VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs  
25 from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer  
30 search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum} \{ \text{length}(\text{Seq. 1}), \text{length}(\text{Seq. 2}) \}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding PKIN are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding PKIN. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

#### VIII. Extension of PKIN Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate

fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ , and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100  $\mu$ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5  $\mu$ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5  $\mu$ l to 10  $\mu$ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector



(Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

5           The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA  
10 recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

          In like manner, full length polynucleotide sequences are verified using the above procedure or  
15 are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

#### **IX. Labeling and Use of Individual Hybridization Probes**

          Hybridization probes derived from SEQ ID NO:23-44 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base  
20 pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 µCi of [ $\gamma$ -<sup>32</sup>P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a  
25 SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10<sup>7</sup> counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

          The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon  
30 membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and

compared.

#### **X. Microarrays**

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Skena (1999), supra.). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Skena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

#### **25 Tissue or Cell Sample Preparation**

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one

with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated  
5 using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

### Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element  
10 is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope  
15 slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

20 Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene).  
25 Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

### Hybridization

30 Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly

larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140  $\mu$ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

## 5 Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide  
10 containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477,  
15 Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

20 The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore,  
25 are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC  
30 computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission

spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used  
5 for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

#### **XI. Complementary Polynucleotides**

Sequences complementary to the PKIN-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PKIN. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with  
10 smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of PKIN. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the PKIN-encoding transcript.

#### **15 XII. Expression of PKIN**

Expression and purification of PKIN is achieved using bacterial or virus-based expression systems. For expression of PKIN in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid  
20 promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express PKIN upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of PKIN in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly  
25 known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PKIN by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the  
30 latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, PKIN is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step,

affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from PKIN at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified PKIN obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII, where applicable.

### XIII. Functional Assays

PKIN function is assessed by expressing the sequences encoding PKIN at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of PKIN on gene expression can be assessed using highly purified populations

of cells transfected with sequences encoding PKIN and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding PKIN and other genes of interest can be analyzed by northern analysis or microarray techniques.

#### XIV. Production of PKIN Specific Antibodies

PKIN substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the PKIN amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, *supra*, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, *supra*.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for anti-peptide and anti-PKIN activity by, for example, binding the peptide or PKIN to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### XV. Purification of Naturally Occurring PKIN Using Specific Antibodies

Naturally occurring or recombinant PKIN is substantially purified by immunoaffinity chromatography using antibodies specific for PKIN. An immunoaffinity column is constructed by covalently coupling anti-PKIN antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing PKIN are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PKIN (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt

antibody/PKIN binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PKIN is collected.

#### XVI. Identification of Molecules Which Interact with PKIN

PKIN, or biologically active fragments thereof, are labeled with  $^{125}\text{I}$  Bolton-Hunter reagent.

5 (See, e.g., Bolton, A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled PKIN, washed, and any wells with labeled PKIN complex are assayed. Data obtained using different concentrations of PKIN are used to calculate values for the number, affinity, and association of PKIN with the candidate molecules.

10 Alternatively, molecules interacting with PKIN are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) *Nature* 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

PKIN may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions  
15 between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

#### XVII. Demonstration of PKIN Activity

Generally, protein kinase activity is measured by quantifying the phosphorylation of a protein  
20 substrate by PKIN in the presence of gamma-labeled  $^{32}\text{P}$ -ATP. PKIN is incubated with the protein substrate,  $^{32}\text{P}$ -ATP, and an appropriate kinase buffer. The  $^{32}\text{P}$  incorporated into the substrate is separated from free  $^{32}\text{P}$ -ATP by electrophoresis and the incorporated  $^{32}\text{P}$  is counted using a radioisotope counter. The amount of incorporated  $^{32}\text{P}$  is proportional to the activity of PKIN. A determination of the specific amino acid residue phosphorylated is made by phosphoamino acid  
25 analysis of the hydrolyzed protein.

In one alternative, protein kinase activity is measured by quantifying the transfer of gamma phosphate from adenosine triphosphate (ATP) to a serine, threonine or tyrosine residue in a protein substrate. The reaction occurs between a protein kinase sample with a biotinylated peptide substrate and gamma  $^{32}\text{P}$ -ATP. Following the reaction, free avidin in solution is added for binding to the  
30 biotinylated  $^{32}\text{P}$ -peptide product. The binding sample then undergoes a centrifugal ultrafiltration process with a membrane which will retain the product-avidin complex and allow passage of free gamma  $^{32}\text{P}$ -ATP. The reservoir of the centrifuged unit containing the  $^{32}\text{P}$ -peptide product as retentate is then counted in a scintillation counter. This procedure allows assay of any type of protein kinase



sample, depending on the peptide substrate and kinase reaction buffer selected. This assay is provided in kit form (ASUA, Affinity Ultrafiltration Separation Assay, Transbio Corporation, Baltimore MD, U.S. Patent No. 5,869,275). Suggested substrates and their respective enzymes are as follows:

- Histone H1 (Sigma) and p34<sup>cdc2</sup>kinase, Annexin I, Angiotensin (Sigma) and EGF receptor kinase,  
5 Annexin II and *src* kinase, ERK1 & ERK2 substrates and MEK, and myelin basic protein and ERK  
(Pearson, J.D. et al. (1991) Methods in Enzymology 200:62-81).

In another alternative, protein kinase activity of PKIN is demonstrated *in vitro* in an assay containing PKIN, 50µl of kinase buffer, 1µg substrate, such as myelin basic protein (MBP) or synthetic peptide substrates, 1 mM DTT, 10 µg ATP, and 0.5µCi [γ-<sup>33</sup>P]ATP. The reaction is  
10 incubated at 30°C for 30 minutes and stopped by pipetting onto P81 paper. The unincorporated [γ-<sup>33</sup>P]ATP is removed by washing and the incorporated radioactivity is measured using a radioactivity scintillation counter. Alternatively, the reaction is stopped by heating to 100°C in the presence of SDS loading buffer and visualized on a 12% SDS polyacrylamide gel by autoradiography. Incorporated radioactivity is corrected for reactions carried out in the absence of PKIN or in the presence of the  
15 inactive kinase, K38A.

In yet another alternative, adenylate kinase or guanylate kinase activity may be measured by the incorporation of <sup>32</sup>P from gamma-labeled <sup>32</sup>P -ATP into ADP or GDP using a gamma radioisotope counter. The enzyme, in a kinase buffer, is incubated together with the appropriate nucleotide mono-phosphate substrate (AMP or GMP) and <sup>32</sup>P-labeled ATP as the phosphate donor. The  
20 reaction is incubated at 37°C and terminated by addition of trichloroacetic acid. The acid extract is neutralized and subjected to gel electrophoresis to separate the mono-, di-, and triphosphonucleotide fractions. The diphosphonucleotide fraction is cut out and counted. The radioactivity recovered is proportional to the enzyme activity.

In yet another alternative, other assays for PKIN include scintillation proximity assays (SPA),  
25 scintillation plate technology and filter binding assays. Useful substrates include recombinant proteins tagged with glutathione transferase, or synthetic peptide substrates tagged with biotin. Inhibitors of PKIN activity, such as small organic molecules, proteins or peptides, may be identified by such assays.

Kinase activity of PKIN may be determined by its ability to convert polyphosphate substrate (PolyP) to ATP in the presence of ADP. PKIN and Poly P are incubated at 37°C for 40 minutes and  
30 then at 90°C for 2 minutes in a buffer containing 50 mM Tris-HCl, pH 7.4, 40 mM ammonium sulfate, 4 mM MgCl<sub>2</sub>, and 5 µM ADP. The reaction mixture is diluted 1:100 in 100 mM Tris-HCl (pH 8.0), 4 mM EDTA, which is then diluted 1:1 in luciferase reaction mixture (ATP Bioluminescence Assay Kit CLS II; Boehringer Mannheim). The ATP generated is then quantitated using a luminometer

(Kornberg, A. et al. (1999) *Annu. Rev. Biochem.* 68:89-125; Ault-Riché, D. et al. (1998) *J. Bacteriol.* 180:1841-1847).

Kinase activity of PKIN, as measured by phosphorylation of substrate, may be determined using an immune complex kinase assay well known in the art. COS7 cells are transfected with an expression plasmid constructed from a FLAG tag expression vector (pME18S-FLAG) containing PKIN DNA. A control transfection using vector alone without the PKIN DNA insert is done in parallel. After 48 hours, the cells are lysed in buffer A (20 mM HEPES-NaOH, pH 7.5, 3 mM MgCl<sub>2</sub>, 100 mM NaCl, 1 mM dithiothreitol, 1 mM phenylmethanesulfonyl fluoride, 1 µg/ml leupeptin, 1 mM EGTA, 1 mM Na<sub>3</sub>VO<sub>4</sub>, 10 mM NaF, 20 mM β-glycerophosphate, and 0.5% Triton X-100) and centrifuged at 14,000 rpm. Supernatants are incubated with anti-FLAG antibody (M2 monoclonal antibody; Eastman Kodak Co.) in a 50% slurry of protein A-Sepharose (Amersham Pharmacia Biotech) for 1.5 hours at 4°C. Immune complexes are precipitated and washed twice in buffer A and twice in buffer B (20 mM HEPES-NaOH, pH 7.5, 1 mM dithiothreitol, 10 µM Na<sub>3</sub>VO<sub>4</sub>, 2 mM β-glycerophosphate, 0.1 mM phenylmethanesulfonyl fluoride, 0.1 µg/ml leupeptin, 0.1 mM EGTA.) Precipitates are incubated in buffer B containing 0.17 mg/ml myelin basic protein (MBP) (Sigma), 20 µM ATP, and 5 µCi of [γ-<sup>32</sup>P]ATP (NEN Life Science Products) at 30°C for 20 minutes. The reaction is stopped by the addition of 4X Laemmli sample buffer (50 mM Tris-HCl, pH 6.8, 2% SDS, 30 mM dithiothreitol, and 10% glycerol) and heated at 95°C for 5 minutes. Proteins are separated by SDS-polyacrylamide gel electrophoresis and radioactivity incorporated into MBP is detected by autoradiography (Nakano, K. et al. (2000) *J. Biol. Chem.* 275:20533-20539.)

In yet another alternative, an assay for PanK activity of PKIN includes the enzyme preparation method as described in Vallari, D.S. et al., (1987) *J. Biol. Chem.* 262:2468-247. Pantothenate kinase-specific activities in cell lysates are calculated as a function of protein concentration with the assay being linear with respect to both time and protein input. Protein concentrations are measured using the Bradford assay using bovine γ-globulin as a standard. Standard assays contain D-[1-<sup>14</sup>C]pantothenate (45.5 µM; specific activity 55 mCi/mmol), ATP (2.5 mM, pH 7.0), MgCl<sub>2</sub> (2.5 mM), Tris-HCl (0.1 M, pH 7.5), and 15 µg of protein from a soluble cell extract in a total volume of 40 µl. The mixture is incubated for 10 min. at 37 °C, and the reaction is stopped by depositing a 30-µl aliquot onto a Whatman DE81 ion-exchange filter disc which is then washed in three changes of 1% acetic acid in 95% ethanol (25 ml/disc) to remove unreacted pantothenate. 4'-Phosphopantothenate is quantitated by counting the dried disc in 3 ml of scintillation solution (Rock, *supra*).

#### XVIII. Enhancement/Inhibition of Protein Kinase Activity

Agonists or antagonists of PKIN activation or inhibition may be tested using assays described in section XVII. Agonists cause an increase in PKIN activity and antagonists cause a decrease in PKIN activity.

- 5           Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious  
10 to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
7482896	1	7482896CD1	23	7482896CB1
7483046	2	7483046CD1	24	7483046CB1
71636374	3	71636374CD1	25	71636374CB1
7480597	4	7480597CD1	26	7480597CB1
3227248	5	3227248CD1	27	3227248CB1
4207273	6	4207273CD1	28	4207273CB1
7483334	7	7483334CD1	29	7483334CB1
7483337	8	7483337CD1	30	7483337CB1
6035509	9	6035509CD1	31	6035509CB1
7373485	10	7373485CD1	32	7373485CB1
5734965	11	5734965CD1	33	5734965CB1
7473788	12	7473788CD1	34	7473788CB1
3107989	13	3107989CD1	35	3107989CB1
7482887	14	7482887CD1	36	7482887CB1
2963414	15	2963414CD1	37	2963414CB1
7477139	16	7477139CD1	38	7477139CB1
55009053	17	55009053CD1	39	55009053CB1
7474648	18	7474648CD1	40	7474648CB1
7483053	19	7483053CD1	41	7483053CB1
7483117	20	7483117CD1	42	7483117CB1
7484498	21	7484498CD1	43	7484498CB1
7638121	22	7638121CD1	44	7638121CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank NO:	ID	Probability Score	GenBank Homolog
1	7482896CD1	g852055		2.90E-167	[Homo sapiens] casein kinase I-alpha Fish, K.J. et al., (1995) J. Biol. Chem. 270:14875-14883
2	7483046CD1	g2736151		4.20E-167	[Rattus norvegicus] myotonic dystrophy kinase-related Leung, T. et al., (1998) Mol. Cell. Biol. 18:130-140
3	71636374CD1	g7549223		0	[Mus musculus] PALS1 (proteins associated with Lin-7, a membrane-associated guanylate kinase) Kamberov, E. et al., (2000) J. Biol. Chem. 275:11425-11431
4	7480597CD1	g2224679		1.40E-97	[Homo sapiens] KIAA0369 doublecortin-like kinase Nagase, T. et al., (1997) DNA Res. 4:141-150
5	3227248CD1	g6690020		4.90E-199	Burgess, H.A. et al. (1999) J. Neurosci. Res. 58:567-575
6	4207273CD1	g4028547		4.70E-68	[Mus musculus] pantothenate kinase 1 beta Rock, C.O. et al. (2000) J. Biol. Chem. 275:1377-1383
7	7483334CD1	g479173		1.70E-251	[Dictyostelium discoideum] MEK kinase alpha Chung, C.Y. et al. (1998) Genes Dev. 12:3564-3578
8	7483337CD1	g9280288		3.10E-27	[Homo sapiens] protein kinase Schultz, S.J. et al. (1994) Cell Growth Differ. 5:625-635
9	6035509CD1	g6110362		3.60E-76	[Arabidopsis thaliana] receptor protein kinase Kaneko, T. et al. (2000) DNA Res. 7:217-221
10	7373485CD1	g4200446		0	[Homo sapiens] Traf2 and NCK interacting kinase, splice variant 7 Fu, C.A. et al. (1999) J. Biol. Chem. 274:30729-30737
11	5734965CD1	g2905643		4.60E-109	[Mus musculus] FYVE finger-containing phosphoinositide kinase Shisheva, A. et al. (1999) Mol. Cell. Biol. 19:623-634
12	7473788CD1	g7160989		1.70E-148	[Klebsiella pneumoniae] ribitol kinase Heuel, H. et al. (1998) Microbiology 144(Pt 6):1631-9
13	3107989CD1	g6690020		1.60E-129	[Homo sapiens] serine/threonine protein kinase Ruiz-Perez VL, et al. (2000) Nat. Genet. 24(3):283-6
					[Mus musculus] pantothenate kinase 1 beta Rock, C.O. et al. (2000) J. Biol. Chem. 275:1377-1383

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank NO:	ID	Probability Score	GenBank Homolog
14	7482887CD1	g205662		3.90E-48	[Rattus norvegicus] nucleoside diphosphate kinase Kimura, N. et al. J. Biol. Chem. (1990) 265:15744-15749
15	2963414CD1	g6524024		8.90E-106	[Mus musculus] mammalian inositol hexakisphosphate kinase I Saiardi, A. et al. Curr. Biol. (1999) 9:1323-1326
16	7477139CD1	g6472874		0	[Mus musculus] Nck-interacting kinase-like embryo specific kinase Nakano, K. et al. J. Biol. Chem. (2000) 275:20533-20539
17	5509053CD1	g15131540		0	[fl][Homo sapiens] (AJ316534) serine/threonine protein kinase
18	7474648CD1	g14346040		0	[fl][Homo sapiens] serine/threonine kinase PSKH2
19	7483053CD1	g5419753		0	[Homo sapiens] RET tyrosine kinase receptor Bordeaux, M.C. et al. (2000) EMBO J. 19:4056-4063
20	7483117CD1	g644770		2.70E-136	[Xenopus laevis] Wee1A kinase Mueller, P.R. et al. (1995) Mol. Biol. Cell 6:119-134
21	7484498CD1	g3599509		0	[Mus musculus] rho/rac-interacting citron kinase Di Cunto, F. et al. (1998) J. Biol. Chem. 273:29706-29711
22	7638121CD1	g212661		1.20E-60	[Gallus gallus] smooth muscle myosin light chain kinase precursor Olson, N.J. et al. (1990) Proc. Natl. Acad. Sci. U.S.A. 87:2284-2288

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7482896CD1	337	S105 S122 S199 S237 S242 S27 S49 S7 S96 T109 T146 T184 T228 T243 T323 T327 T38 Y209 Y274	N167 N215 N3	Eukaryotic protein kinase domain: Y17-F211 Protein kinases signatures and profile: T112-R168 PROTEIN KINASE DOMAIN DM00004 P35506 19-273: L19-Y274 P54367 22-276: L19-Y274 P48730 11-265: L19-Y274 B56406 19-273: L19-Y274 CASEIN KINASE I TRANSFERASE SERINE/THREONINE PROTEIN ATP-BINDING ISOFORM ALPHA CKI ALPHA MULTIGENE PD006522: R282-G324 Tyrosine kinase catalytic domain PR00109: Y126-M144 Kinase Protein Domain PD00584: V20-G29 Protein kinases ATP-binding region signature: I23-K46 Serine/Threonine protein kinases active-site signature: F132-M144 signal_cleavage: M1-G40	HMMER-PFAM PROFILE-SCAN BLAST-DOMO BLAST-PRODROM BLIMPS-PRINTS BLIMPS-PRODROM MOTIFS MOTIFS SPSCAN HMMER-PFAM
2	7483046CD1	475	S161 S280 S307 S363 S407 S430 T455		Eukaryotic protein kinase domain: F71-F337	

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PROTEIN KINASE DOMAIN DM00004  Q09013 83-336: I73-R325 S42867 75-498: I73-H252 I38133 90-369: E72-L220 P53894 353-658: L74-G215 KINASE PHORBOLESTER BINDING DYSTROPHY KINASE RELATED CDC42 BINDING SIMILAR SERINE/THREONINE PROTEIN GENGHIS KHAN PD012280: L25-D70 Tyrosine kinase catalytic domain PR00109: M148-S161, S185-L203, C257-E279 Protein kinase C terminal domain: P351-D366 Protein kinases ATP-binding region signature: I77-K100 Serine/Threonine protein kinases active-site signature: Y191-L203 signal_cleavage: M1-S37 Guanylate kinase: T515-I624	BLAST-DOMO
3	71636374CD1	675	S130 S14 S143 S25 S383 S432 S517 S562 S569 S576 S581 S646 S84 T137 T253 T270 T422 T465 T514 T558 T584 T97 Y593	N82	GUANYLATE KINASE DM00755 A57653 370-570: P475-P670 P54936 769-955: R478-P670 I38757 709-898: Q474-P670 P31016 529-718: R480-P670	BLAST-DOMO



Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PROTEIN DOMAIN SH3 KINASE GUANYLATE TRANSFERASE ATPBINDING REPEAT GMP MEMBRANE PD001338: T514-E620	BLAST- PRODROM
					SIMILAR TO GUANYLATE KINASE PD065809: G41-Q337	BLAST- PRODROM
					Guanylate kinase protein BL00856: V511-V531, D539-R586	BLIMPS- BLOCKS
					SH3 domain signature PR00452: D386-E395, I348-P358, L369-Q384	BLIMPS- PRINTS
					PDZ domain (Also known as DHR or GLGF). PDZ: I256-S335,	HMMER- PFAM
					SH3 domain SH3:I348-Q415	HMMER- PFAM
					ATP/GTP-binding site motif A (P-loop): A404-S411	MOTIFS
					Guanylate_Kinase signature and profile: T514-V531	MOTIFS
4	7480597CD1	835	S11 S153 S174 S223 S249 S271 S292 S349 S369 S380 S389 S393 S405 S525 S54 S59 S633 S713 T129 T194 T246 T278 T300 T319 T33 T451 T477 T499 T514 T545 T610 T63 T681 T790 T808	N768	Eukaryotic protein kinase domain pkinase: Y543-I800	HMMER- PFAM
					Protein kinases signatures and profile: D640-I697	PROFILE- SCAN
					PROTEIN KINASE DOMAIN DM00004 S57347 21-266: V548-T790 P08414 44-285: I549-T790 A44412 16-262: I549-A791 JU0270 16-262: I549-A791	BLAST- DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					KINASE PROTEIN TRANSFERASE ATP-BINDING SERINE/THREONINE PROTEIN PHOSPHORYLATION RECEPTOR TYROSINE PROTEIN PRECURSOR TRANSMEMBRANE PD000001: E609-V693	BLAST-PRODROM
					Octicosapeptide repeat p PF00564: Y543-S597, H605-M655, K473-G526	BLIMPS-PFAM
					Tyrosine kinase catalytic domain PR00109: L618-I631, H654-V672	BLIMPS-PRINTS
					Protein kinases ATP-binding region signature: I549-K572	MOTIFS
					Serine/Threonine protein kinases active-site signature: I660-V672	MOTIFS
5	3227248CDI	373	S100 S283 S285 S330 S47 T10 T167 T209 T226 T230 T244 T34	N103 N72		
6	4207273CDI	735	S100 S111 S113 S124 S152 S170 S179 S185 S186 S20 S202 S215 S221 S225 S240 S267 S271 S302 S459 S503 S729 S9 T10 T105 T13 T30 T402 T417 T425 T469 T626 T663 T669 T84 Y512	N289 N312 N341 N392 N400 N61 N624 N647	PROTEIN KINASE DOMAIN DM00004 A48084 98-348:K470-A722 DM00004 Q01389 1176-1430:K470-A722 DM00004 P41892 11-249:G471-R719 DM00004 Q10407 826-1084:K470-A722	BLAST-DOMO
					KINASE PROTEIN TRANSFERASE ATP-BINDING SERINE/THREONINE PROTEIN PHOSPHORYLATION RECEPTOR TYROSINE PROTEIN PRECURSOR TRANSMEMBRANE PD000001:L631-P673, E472-C537, Y533-S633, S701-S734	BLAST-PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	7483334CD1	506	S148 S206 S243 S319 S325 S354 S47 T197 T288 T293 T308 T321 T373 T386 T402 T403 T479	N181 N345 N377 N401	Tyrosine kinase catalytic domain signature PR00109:M547-N560, Y583-L601, G636-I646, S655-M677	BLIMPS- PRINTS
					Eukaryotic protein kinase domain: pkinase: W468-L731	HMMER- PFAM
					Protein_Kinase_Atp L474-K496	MOTIFS
					Protein_Kinase_St V589-L601	MOTIFS
					Protein kinases signatures and profile protein_kinase_tyr.prf:V569-A619	PROFILE- SCAN
					PROTEIN KINASE DOMAIN DM00004	BLAST- DOMO
					P51954 6-248:L7-S247	
					P51957 8-251:L7-S247	
					P51955 10-261:V6-S247	
					Q08942 22-269:M9-S247	
8	7483337CD1	2014	S1076 S1151 S1177 S1217 S1274 S1279 S1454 S15 S1515 S1679 S1700 S1811 S1833 S1887 S1890 S1999	N1024 N1119 N1338 N1599 N1674 N307 N371 N409	Tyrosine kinase catalytic domain signature PR00109:M79-K92, H117-L135, S183-N205, Y226-A248	BLIMPS- PRINTS
					Eukaryotic protein kinase domain pkinase: Y4-V257	HMMER- PFAM
					Protein_Kinase_Atp I10-K33	MOTIFS
					Protein_Kinase_St V123-L135	MOTIFS
					Protein kinases signatures and profile protein_kinase_tyr.prf:M103-M156	PROFILE- SCAN
					PROTEIN KINASE DOMAIN DM00004	BLAST- DOMO
					I38044 100-349:I1295-P1549	
					I49663 194-437:E1341-P1549	
					A53800 119-368:R1343-P1549	
					S29851 157-404:E1341-P1549	

Table 3

SEQ ID NO:	Incyle Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	6035509CD1	348	S203 S25 S321 S337 S401 S531 S56 S565 S599 S81 S843 S863 S887 S900 T1091 T1099 T1113 T1187 T1189 T1234 T1401 T1543 T1605 T1634 T1660 T1872 T1895 T2010 T280 T494 T517 T524 T533 T537 T680 T687 T699 T702 T703 T753 T795 T811 T835 T909 Y1225 Y1997 Y907	N177	Tyrosine kinase catalytic domain signature PR00109: Y1414-V1432, V1483-H1505, Q1529-A1551	BLIMPS-PRINTS
			transmembrane domain transmem_domain:P1367-N1387		HMMER	
			Eukaryotic protein kinase domain pkinase:E1280-P1549		HMMER-PFAM	
			Protein kinases signatures and profile protein_kinase_tyr.prf:L1400-E1457		PROFILE-SCAN	
			Atp_Gtp_A G672-S679		MOTIFS	
			PROTEIN KINASE DOMAIN DM00004 P10676 18-272:I17-P270 A53714 17-262:I17-S271 P38692 24-266:E19-S271 P08458 20-262:I21-S271		BLAST-DOMO	
			Tyrosine kinase catalytic domain signature PR00109:H134-L152, G181-I191, W250-V272		BLIMPS-PRINTS	
			Eukaryotic protein kinase domain pkinase:W15-I281		HMMER-PFAM	
			Protein_Kinase_Atp I21-K44		MOTIFS	
			Protein_Kinase_St I140-L152		MOTIFS	
10	7373485CD1	2042	S1020 S105 S1079 S1125 S1130 S1148 S13	N1061 N1274 N1647 N1671	Protein kinases signatures and profile protein_kinase_tyr.prf:M120-T172 Probable phosphatidyl inositol 4-phosphate 5-kinase FAB I EC 2.7.1.68 1-phosphatidyl inositol 4-	PROFILE-SCAN BLAST-PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			S1377 S1419 S1429 S1440 S1466 S1483 S1488 S1544 S1545 S1620 S1639 S1648 S168 S1685 S1703 S1784 S1785 S1830 S1899 S228 S244 S257 S261 S286 S291 S367 S423 S475 S502 S576 S789 S810 S835 S85 S872 S896 S977 T1005 T1013 T109 T1149 T1295 T1386 T1524 T1567 T1670 T1674 T1681 T1708 T1722 T173 T1743 T1813 T1852 T1872 T1909 T1970 T341 T342 T591 T666 T731 T782 T976 T984 Y1290 Y1716Y1933 Y659	N1870 N303 N310 N333	phosphate 5-kinase diphosphoinositide transferase PD136025:H461-F821, W1147-K1437, L1375-S1702, K638-K767, P1663-V1780, D1372-Q1458, F959-I1069, R960-D1053, F200-R262, D1895-S1950; PD041996:L1974-W2035 5-KINASE PHOSPHATIDYL INOSITOL 4-PHOSPHATE KINASE TYPE TRANSFERASE DIPHOSPHOINOSITIDE 1-PHOSPHATIDYL INOSITOL 4-PHOSPHATE II ALPHA PHOSPHATIDYL INOSITOL PD002308:P1751-G1966, L1974-F2028, I493-H533 FYVE zinc finger FYV:Q153-C213 Phosphatidylinositol-4-phosphate 5-Kinase PIP5:R1791-F2028	BLAST- PRODROM
11	5734965CD1	551	S107 S176 S2 S21 S257 S368 S502 S54 T183 T286 T334 T356 T403 T66 Y526 Y531	N127 N219	FGGY family of carbohydrate kinases: L423-A490 FGGY FAMILY OF CARBOHYDRATE KINASES DM01757[P21939][1-480: V13-A184 XYLULOKINASE DM02388[P18157][1-492: T383-E539	HMMER- PFAM BLAST- DOMO BLAST- DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					FGY FAMILY OF CARBOHYDRATE KINASES DM01757 P37677 I-479: R10-D260 FGY FAMILY OF CARBOHYDRATE KINASES DM01757 P46834 I-488: Y11-V268 MPA43 PROTEIN PDI30314: V13-I210	BLAST- DOMO BLAST- DOMO BLAST- PRODOM
					FGY family of carbohydrate kinases proteins BL00933: Y11-L34, R109-A119, V137-N156, G456-I471	BLIMPS- BLOCKS
12	7473788CD1	485	S10 S159 S3 S343 S362 S415 S417 T115 T192 T466 T469 T76 Y119	N405	Eukaryotic protein kinase domain: F93-Q345 PROTEIN KINASE DOMAIN DM00004 P54644 I22-362: I95-S342 PROTEIN KINASE DOMAIN DM00004 P28178 I55-393: I95-L341 PROTEIN KINASE DOMAIN DM08046 P05986 I-397: K65-P372 P06244 I-396: F93-P372 Tyrosine kinase catalytic domain signature PR00109: V170-Q183, Y206-L224 Protein kinases ATP-binding region signature I99-K122 Serine/Threonine protein kinases active-site signature I212-L224 signal_cleavage: M1-A24 signal_cleavage: M1-A27	HMMER- PFAM BLAST- DOMO BLAST- DOMO BLAST- DOMO BLIMPS- PRINTS MOTIFS MOTIFS SFSCAN SFSCAN
13	3107989CD1	282	S148 S152 S192 S194 S239 S78 T118 T138 T139 T153 T36	N12		

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14	7482887CD1	151	S42 S97 T35 Y141		NUCLEOSIDE DIPHOSPHATE KINASES DM00773 P48817 3-152:I7-Y150 DM00773 I39074 19-168:I7-Y150 DM00773 Q07661 1-148:I7-Y150 DM00773 P50590 1-150:I7-Y150 KINASE DIPHOSPHATE NUCLEOSIDE TRANSFERASE NDK NDP ATP-BINDING PROTEIN I PRECURSOR PD001018:I7-Y150 Nucleoside diphosphate kinases proteins BL00469:E77-L131 Nucleoside diphosphate kinases NDK:I7-A151	BLAST- DOMO
					Nucleoside diphosphate kinases active site ndp_kinases:G96-R142	PROFILE- SCAN
15	2963414CD1	410	S134 S156 S276 S318 T259 T361 T374 T383 T62	N117 N290	PROTEIN ARGININE METABOLISM REGULATION III TRANSCRIPTION SIMILARITY SACCCHAROMYCES CEREVISIAE PUTATIVE PD011544:S188-Q333, S355-L403 PUTATIVE BZIP TRANSCRIPTION FACTOR CHROMOSOME IV READING FRAME ORF YDR017C PD024140:G15-R197 Aldo/keto reductase family putative active site signature I312-L327	BLAST- PRODROM  BLAST- PRODROM  MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16	7477139CD1	1581	S101 S1107 S1112 S1139 S1178 S1233 S1291 S1346 S136 S1400 S1426 S1435 S148 S1537 S1577 S211 S283 S376 S498 S580 S671 S676 S700 S709 S718 S749 S807 S84 S890 S891 S892 S910 T1071 T1123 T1194 T1367 T1508 T1546 T1556 T246 T276 T294 T357 T573 T664 T690 T899 T981 T992	N1137 N1201 N146 N654 N668 N990	PROTEIN KINASE DOMAIN DM00004 P10676 18-272:Y83-P302 DM00004 A53714 17-262:L43-S304 DM00004 P38692 24-266:S84-C293, K29-N57 DM00004 P50527 388-627:K77-S304, I31-E65	BLAST_DO MO
17	55009053CD1	1084	S1024 S1031 S1038 S1042 S1058 S157 S172 S231 S25 S422 S452 S478 S52 S521 S552 S569 S604 S623 S709 S80 S862 S882 S895 S914 S962 S968 S969 S981 S988 T102 T1037 T167 T230 T256 T263 T37 T420 T48 T543 T593 T631 T8 Y1005	N953	Serine/Threonine protein kinases active-site signature I139-I151 Protein kinases signatures and profile protein_kinase_1yr.prf: L118-F173 Eukaryotic protein kinase domain pkinase: L15-F273 Tyrosine kinase catalytic domain PR00109: T95-R108, H133-I151, V197-C219, K242-I264 PROTEIN KINASE DOMAIN DM00004 S49611 39-259: I21-K242 Q05609 553-797: E20-C253 P51957 8-251: I21-R261 P41892 11-249: I21-R261	MOTIFS PROFILE-SCAN HMMER-PFAM BLIMPS-PRINTS BLAST-DOMO
18	7474648CD1	600	S206 S331 S369 S425 S456 S543 S55 S571	N18 N495	Protein kinases ATP-binding region signature I284-K307	MOTIFS



Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			S577 S585 T117 T14 T25 T299 T300 T356 T371 T395 T433 T58		Eukaryotic protein kinase domain pkinase: Y278-V535 Tyrosine kinase catalytic PR00109: M352-I365, Y388-Y406, V458-E480 PROTEIN KINASE DOMAIN DM00004 S57347 21-266: D279-L516 P08414 44-285: I280-S525 JN0323 25-268: I284-R523 S46284 28-274: I284-A526	HMMER-PFAM BLIMPS-PRINTS BLAST-DOMO
19	7483053CD1	1114	S1034 S104 S110 S131 S159 S173 S224 S363 S413 S457 S522 S561 S65 S670 S691 S696 S765 S811 S819 S836 S922 T1022 T1055 T1078 T261 T295 T315 T328 T350 T456 T492 T538 T564 T675 T729 T75 T847 T930 Y1096 Y483 Y905	N1092 N151 N199 N336 N343 N361 N367 N377 N394 N448 N468 N554 N834 N975 N98	signal peptide: M1-G28 Signal_cleavage: M1-A26 Transmembrane domain: L13-F31 Protein kinases ATP-binding region signature L730-K758 Tyrosine protein kinases specific active-site signature L870-V882 Protein kinases signatures and profile protein_kinase_tyr.prf: D850-D903 Receptor tyrosine kinase class II signature receptor_tyr_kin_ii.prf: R878-D925 Cadherin domain cadherin: P172-T261	HMMER SPSCAN HMMER MOTIFS MOTIFS PROFILE-SCAN PROFILE-SCAN HMMER-PFAM HMMER-PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Receptor tyrosine kinase BL00239: D903-Y952, P957-I1001, E775-V822, L851-R873, A876-E901 BL00240: K716-A764, A764-E818, D850-K887, E902-G949, G949-I1001 BL00790: G748-L801, A855-A876, A877-D903, Q910-W942, H968-L1016	BLIMPS-BLOCKS
					Tyrosine kinase catalytic PR00109: V804-R817, Y864-V882, I913-L923, S932-G954, C976-F998 RECEPTOR KINASE PRECURSOR SIGNAL RET TYROSINE PROTOONCOGENE TYROSINE CRET TRANSFERASE PD014372: P273-K666, D300-V725; PD014143: Y30-C197; PD007958: V1010-G1063, PD010335:M1064-S1114	BLIMPS-PRINTS BLAST-PRODROM
					PROTEIN-TYROSINE KINASE RET DM05080 P07949 302-723: D302-L724 I48735 303-724: D302-L724 PROTEIN KINASE DOMAIN DM00004 JN0290 88-360: V725-F998 P07949 725-997: V725-F998	BLAST-DOMO
20	7483117CD1	567	S162 S17 S206 S243 S278 S543 S552 S70 T112 T125 T22 T246 T544 T559 T68 Y238	N15 N332	Protein kinases ATP-binding region signature I218-K241 Serine/Threonine protein kinases active-site signature M335-I347 Eukaryotic protein kinase domain pkinase: F212-L480	MOTIFS MOTIFS HMMER-PFAM
					Tyrosine kinase catalytic site PR00109: N289-S302, Y329-I347, A415-G437, L455-A477	BLIMPS-PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					WEE1 HOMOLOG WEE1 LIKE PROTEIN KINASE MITOSIS TRANSFERASE TYROSINE PROTEIN ATP BINDING PHOSPHORYLATION PD028078: N483-G561 PROTEIN KINASE DOMAIN DM00004 P47817211-470: L213-A477 P30291300-559: E214-A477 P54350241-507: E214-A477 A57247104-343: K217-I347, A366-R474	BLAST- PRODOM
21	7484498CD1	2054	S81 S93 S140 S248 S308 S361 S381 S386 S410 S436 S445 S480 S487 S501 S516 S529 S546 S577 S582 S699 S883 S888 S924 S1031 S1049 S1097 S1158 S1160 S1234 S1315 S1364 S1365 S1370 S1371 S1377 S1574 S1845 S1915 S1933 S2014 S2028 T83 T378 T498 T604 T840 T951 T956 T989 T1041 T1062 T1112 T1186 T1231 T1309 T1326 T1336 T1372 T1543 T1583	N835 N1622 N1745 N1768	CNH (NIK-1 like kinase) domain: L1619-Y1916 Phorbol esters/diacylglycerol binding: H1390-C1438 PH (pleckstrin homology) domain: L1471-A1590 Eukaryotic protein kinase domain: F97-F360 Phorbol esters / diacylglycerol binding domain dag_pe_binding_domain.prf: C1403-E1466 Tyrosine kinase catalytic domain signature PR00109: S211-V229, C284-G306, M174-N187 Domain found in NIK1-like kinase, mouse citron and yeast ROM1, ROM2 PF00780: K534-I542, N891- T933, I964-Q975, Q1015-Q1067, Q1217-E1255, I1388-L1434, E1759-A1802, N1819-F1831, K1851- Q1880	HMMER- PFAM HMMER- PFAM HMMER- PFAM HMMER- PFAM PROFILE- SCAN BLIMPS- PRINTS BLIMPS- PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			T1775 T1787 T1943 T1955 T1961 T2015 Y763		CITRON PROTEIN COILED COIL RHO/RACINTERACTING KINASE PD155701: F859-L1071 PD143273: G1439-V1631 PD082663: L1201-P1389 PD143272: A1881-V2054 PROTEIN KINASE DOMAIN DM00004 Q09013 83-336: V99-L349 S42867 75-498: S101-G241, I258-S445 S42864 41-325: E98-G241, N249-L349 P53894 353-658: L102-G241 I258-L349 Protein kinases ATP-binding region signature V103-K126	BLAST- PRODROM
					Serine/Threonine protein kinases active-site signature: Y217-V229 Leucine zipper pattern: L854-L875, L991-L1012, L1057-L1078, L1159-L1180 Carbamoyl-phosphate synthase subdomain signature 2: M1172-S1179 Phorbol esters / diacylglycerol binding domain: H1390-C1438	MOTIFS
					Immunoglobulin domain: G68-A128, G1174-V1235	MOTIFS
22	7638121CD1	1665	S97 S152 S156 S163 S242 S364 S450 S459 S491 S493 S528 S536 S588 S762 S827 S875 S915 S917 S929 S947 S961 S997 S1087 S1147	N1005	Eukaryotic protein kinase domain: Y165-F418, F1369-L1621 Protein kinases signatures and profile protein_kinase_tyr.prf: E260-A314	HMMER- PFAM HMMER- PFAM PROFILE- SCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			S1203 S1336 S1351 S1365 S1391 S1434 S1446 S1459 S1461 S1521 T59 T230 T257 T312 T668 T870 T966 T1211 T1310 T1320 T1638		Tyrosine kinase catalytic domain signature PR00109: S341-E363, L387-A409, L238-Y251, Y274-M292	BLIMPS-PRINTS
					KINASE PROTEIN TRANSFERASE ATPBINDING SERINE/THREONINEPROTEIN PHOSPHORYLATION RECEPTOR TYROSINEPROTEIN PRECURSOR TRANSMEMBRANE PD000001: V256-V327, S323-D365, S380-P423	BLAST-PRODOR
					PROTEIN KINASE DOMAIN DM00004 JN0583/727-969: V167-R401, Q1372-P1563 P07313/298-541: K168-A409, Q1378-P1563 P53355/15-257: E169-R406, Q1374-P1563 S07571/5152-5396: E166-R406, Q1374-P1606	BLAST-DOMO
					Protein kinases ATP-binding region signature I171-K194	MOTIFS
					Tyrosine protein kinases specific active-site signature I1484-I1496	MOTIFS
					Protein kinase St V280-M192	MOTIFS

Table 4

Polynucleotide SEQ ID NO:	Incyte ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
23	7482896CB1	1014	982-1014	GNN.g7899226_000043_002 .edit	1	1014
24	7483046CB1	1530	719-770, 1-61, 1036-1104, 1271-1461, 313-464	71583296V1 71581650V1 71601507V1 55143579J1 71579961V1 55140831J1	889 778 1124 1 266 118	1476 1455 1530 272 884 522
25	71636374CB1	3150	1294-1806, 1-115, 2593-2616	183812R7 (CARDNOT01) 7676860H1 (NOSETUE01) 8252304H1 (BRAHDIT10) 5223511F9 (OVARDT07) GBL.g7452884_edit GBL.g8919852_edit 7214961H1 (LUNGFEC01) 7710619J1 (TESTTUE02) 7391509H1 (LIVRFEE02) 5958404H1 (BRATNOT05) 5971916H1 (BRAZNOT01)	2581 250 25 1225 1125 1099 1 1611 751 2796 2211	3148 864 804 1397 2085 1898 250 2273 1302 3150 2832
26	7480597CB1	2901	1907-1981, 1-156, 748-1606, 255-313	55150024J1 55073631J1 55150108J1 2841337T6 (DRGLNOT01) 55144761T1 5543295F7 (TESTNOC01) GNN.g7658410_000016_002 56001404J1	1377 630 1711 2251 2132 137 1 1790	2056 1518 2070 2901 2833 574 2013 2434

Table 4

Polynucleotide SEQ ID NO:	Incye ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
27	3227248CB1	1671	1-85, 1593-1671, 1327-1360	70944845V1 7207691H1 (FIBPFEA01) 8283762T1 (LIVRNON08) GB1.g9796547_edit 71281138V1 5260904F6 (CONDTUT01) 5543515F6 (TESTNOC01) 5357164H1 (TESTNOC01) 55144823H1 GNN.g9230839_000001_002	997 451 180 1 1089 569 907 238 2112 1	1646 1050 562 1539 1671 1065 1376 440 2577 1293
28	4207273CB1	2577	1-1641, 1845-1889	55073166J1 4919885T6 (TESTNOT11) 71341632V1 71341335V1 940589R6 (ADRENOT03) 6512850H1 (THYMDIT01) 6102073H1 (UTRENOT09) 4970029F7 (KIDEUNC10) 7659406H1 (OVARNOE02)	1115 1445 1559 1145 1916 1007 797 1 509	1773 2141 2110 1708 2110 1688 1087 677 1081
29	7483334CB1	2110	1-640, 1255-1314, 948-1005	7383958R8 (FTUBTUE01) 3245584H1 (BRAINT019) 72334852V1 7383958F8 (FTUBTUE01) 58002303T1 70771904V1 GNN.g6693375_000016_002 edit 55046508H1	1 2681 5219 537 6221 5851 986 2906	694 2928 5761 1196 7093 6475 3303 3666
30	7483337CB1	7093	1-3002, 4789-5840, 7069-7093, 3561-3671			

Table 4

Polynucleotide SEQ ID NO:	Incye ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
				55144427J1	5514	6397
				5208289H1 (BRAFN02)	4900	5138
				7036825F6 (UTRSTMR02)	3953	4647
				55046508J1	3448	4132
				70772942V1	5079	5680
				6436908H1 (LUNGNON07)	908	1407
				GNN.g6721428_000012_004 .edit	3780	6267
31	6035509CB1	1800	152-333, 1-25, 1463-1800, 770-862	71927475V1	1340	1800
				6035509F8 (PITUNOT06)	848	1614
				55071284J1	818	1098
				72420180D1	1	729
				55071288J1	480	1096
32	7373485CB1	6347	4445-5413, 728-786, 6321-6347, 1497-3441, 4019-4079, 877-1082	72375809V1	2075	2717
				8116978H1 (TONSDIC01)	1	659
				GNN.g6114949_010.edit5p	1497	3728
				6919538R8 (PLACFER06)	1156	1644
				GNN.g6850654_000027_002	998	1496
				7368965H1 (ADREFEC01)	5742	6347
				6460173H2 (OSTEUNC01)	5357	5883
				6801172F6 (COLENOR03)	4290	4817
				7212618T8 (LUNGFEC01)	3001	3712
				6919538F8 (PLACFER06)	390	1143
				55073317H1	2592	3387
				58003367H1	4871	5725
				7271932R8 (OVARDIJ01)	3542	4220
				5623962R8 (THYMNOR02)	4544	5050



Table 4

Polynucleotide SEQ ID NO:	Incyte ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
				72373545V1	1602	2203
				5623962F8 (THYMNOR02)	3970	4319
33	5734965CB1	1876	1-902	3254961T6 (OVARTUN01)	1276	1876
				5897065H1 (BRAYDIN03)	1	291
				70810516V1	181	806
				70162895V1	1002	1658
				70809778V1	915	1490
				70807962V1	302	989
34	7473788CB1	1487	1-121, 1450-1487	70995937V1	1024	1487
				7177378H1 (BRAXDIC01)	29	554
				GNN:g3983531_000002_00 2.edit.1	1	260
				70996158V1	594	1243
				7177563H2 (BRAXDIC01)	489	1180
35	3107989CB1	1884	1-306, 1253-1884	70942785V1	1153	1507
				3107989F6 (BRSTTUT15)	232	609
				7363877H1 (OVARDIC01)	1358	1884
				GNN:g9368012.edit1	375	1465
				2243506F6 (PANCUTUT02)	1	385
36	7482887CB1	1070	1-660, 891-948	56009164H1	1	725
				GBL:g5815507.edit	612	997
				GBL:g9716284_order_0.edit2	988	1070
37	2963414CB1	2890	1-270, 1973-2064, 2658-2890, 726-1584	71883559V1	470	1087
				6741017F6 (BRAFDIT02)	1687	2299
				72524920V1	984	1725
				7090654H1 (BRAUTDR03)	2284	2876

Table 4

Polynucleotide SEQ ID NO:	Incye ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
				7595015H1 (LJVRNOC07)	1	450
				71882107V1	424	985
				70523289V1	1123	1749
				7236935H1 (BRAINYO2)	1904	2302
				2601508H1 (UTRSNOT10)	2660	2890
38	7477139CB1	5198	2528-2698, 1296-2145, 2792-4455, 528-724, 177-214	GNN.g1149521_002	948	3957
				71143326V1	4891	5198
				55117016H1	1	919
				2879284F6 (UTRSTUT05)	4388	4874
				3900926H1 (LUNGNON03)	3689	3971
				GNN.g2780172_002.edit	3433	4943
				72615067V1	701	1315
				6775332H1 (OVARDIR01)	4605	5193
				7369832H1 (ADREFEC01)	4063	4606
39	55009053CB1	3969	1393-2860, 1-649	8036923H1 (SMCRUNE01)	1289	2065
				72480126D1	3325	3969
				7263320F6 (PROSTMC02)	1510	2343
				55009061H1	570	1318
				72476437D1	3306	3968
				6583144F8 (BRAVTXC01)	1	452
				72508467V1	2287	3200
				72509180V1	2494	3329
				55009045J1	288	982
40	7474648CB1	1803	198-1803	FL7474648_g7596812_0000	823	1497
				12_g7981277_1_1		
				GNN.g7596812_2	1	1803
41	7483053CB1	3472	1-305, 3134-3472	GBI.g6981824_000001.edit	1	337
				2493520F6 (ADRETUT05)	2055	2525

Table 4

Polynucleotide SEQ ID NO:	Incye ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
				72498890V1	1524	2231
				GNN.g6981824_000001_042 .edit	74	3187
				55081239H1	847	1704
				6872245H1 (BRAGNON02)	2354	3059
				7995993H1 (ADRETUC01)	2942	3472
				7742567H1 (ADRETUE04)	647	1183
42	7483117CB1	1704	1-342, 509-539, 582-758	GBI.g4153871_000001.edit	1536	1704
				7369322F8 (ADREFEC01)	343	501
43	7484498CB1	6298	4050-4677, 1-195, 623-1785,	GNN.g4153871_006.edit	1	1678
				55058386H1	601	1357
				7073440H1 (BRAUTDR04)	5165	5621
				7032228R8 (BRAXTDR12)	4000	4590
				55053104J1	1618	2321
				7014254F6 (KIDNNOC01)	4579	5133
				7066070H1 (BRATNOR01)	2926	3470
				55053152H1	848	1564
				55058386J1	1	701
				7073642H1 (BRAUTDR04)	5045	5617
				6892089F6 (BRAITDR03)	2294	2708
				8267244H1 (MXDUNF04)	4401	5097
				7076436H1 (BRAUTDR04)	3497	4047
				7068147R8 (BRATNOR01)	5186	5924
				GNN.g4508157_002.edit	1166	1941

Table 4

Polynucleotide SEQ ID NO:	Incye ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
				7741468H1 (THYMNOE01)	3001	3627
				6850478H1 (BRAIFEN08)	5720	6298
				7068147F8 (BRATNOR01)	4092	4592
44	7638121CB1	5454	1718-3145, 1-989, 3982-4016	6736753J1 (SINTFER02)	3907	4637
				7361161H1 (BRAIFEE05)	1	637
				55057003J1	252	937
				56000546J1	1303	2019
				7354408H1 (HEARNON03)	5008	5454
				5863411F6 (MUSLTD01)	3355	4178
				71873215V1	4520	5227
				71875134V1	3114	3669
				6496171T6 (COLNNOT41)	4710	5416
				55141853J2	810	1390
				7647137H1 (UTRSTUE01)	1920	2257
				7600017R6 (ESOGTME01)	1475	2041
				6200811F6 (PITUNON01)	3037	3632
				55052669H1	2245	3081

Table 5

Polynucleotide SEQ ID NO:	Incye Project ID:	Representative Library
24	7483046CB1	COLCTUT03
25	71636374CB1	CARDNOT01
26	7480597CB1	DRGLNOT01
27	3227248CB1	COTRNOT01
28	4207273CB1	TESTNOC01
29	7483334CB1	ADRENOT03
30	7483337CB1	UTRSTMR02
31	6035509CB1	PITUNOT06
32	7373485CB1	MCLDTXT02
33	5734965CB1	PROSTUS23
34	7473788CB1	BRAINOT19
35	3107989CB1	STOMFET02
37	2963414CB1	SCORNOT04
38	7477139CB1	PLACFER06
39	55009053CB1	SINITME01
41	7483053CB1	BRAYDIN03
42	7483117CB1	ADREFEC01
43	7484498CB1	BRAITDR03
44	7638121CB1	MUSLTDR02

Table 6

Library	Vector	Library Description
ADREFEC01	pINCY	This large size-fractionated library was constructed using RNA isolated from adrenal tissue removed from a Caucasian female fetus who died from anencephalus after 16-weeks' gestation. Serology was negative. Family history included taking daily prenatal vitamins and mitral valve prolapse in the mother.
ADRENOT03	PSPORT1	Library was constructed using RNA isolated from the adrenal tissue of a 17-year-old Caucasian male, who died from cerebral anoxia.
BRAINOT19	pINCY	Library was constructed using RNA isolated from diseased brain tissue removed from the left frontal lobe of a 27-year-old Caucasian male during a brain lobectomy. Pathology indicated a focal deep white matter lesion, characterized by marked gliosis, calcifications, and hemosiderin-laden macrophages, consistent with a remote perinatal injury. This tissue also showed mild to moderate generalized gliosis, predominantly subpial and subcortical, consistent with chronic seizure disorder. The left temporal lobe, including the mesial temporal structures, showed focal, marked pyramidal cell loss and gliosis in hippocampal sector CA1, consistent with mesial temporal sclerosis. GFAP was positive for astrocytes. The patient presented with intractable epilepsy, focal epilepsy, hemiplegia, and an unspecified brain injury. Patient history included cerebral palsy, abnormality of gait, and depressive disorder. Family history included brain cancer.
BRAITDR03	PCDNA2.1	This random primed library was constructed using RNA isolated from allocortex, cingulate posterior tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRAYDIN03	pINCY	This normalized library was constructed from 6.7 million independent clones from a brain tissue library. Starting RNA was made from RNA isolated from diseased hypothalamus tissue removed from a 57-year-old Caucasian male who died from a cerebrovascular accident. Patient history included Huntington's disease and emphysema. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 -hours/round) reannealing hybridization was used. The library was linearized and recircularized to select for insert containing clones.
CARDNOT01	PBLUESCRIPT	Library was constructed using RNA isolated from the cardiac muscle of a 65-year-old Caucasian male, who died from a gunshot wound.

Table 6

Library	Vector	Library Description
COLCTUT03	pINCY	Library was constructed using RNA isolated from cecal tumor tissue removed from a 70-year-old Caucasian female during right hemicolectomy, open liver biopsy, flexible sigmoidoscopy, colonoscopy, and permanent colostomy. Pathology indicated invasive grade 2 adenocarcinoma forming an ulcerated mass 2 cm distal to the ileocecal valve and invading the muscularis propria. One regional lymph node (of 16) was positive for metastatic adenocarcinoma. Patient history included a deficiency anemia, malignant breast neoplasm, type II diabetes, hyperlipidemia, viral hepatitis, an unspecified thyroid disorder, osteoarthritis, a malignant skin neoplasm, and normal delivery. Family history included cardiovascular and cerebrovascular disease, hyperlipidemia, and breast and ovarian cancer.
COTRNOT01	pINCY	Library was constructed using RNA isolated from diseased transverse colon tissue obtained from a 26-year-old Caucasian male during a total abdominal colectomy and colostomy. Pathology indicated minimally active pancolitis with areas of focal severe colitis with perforation, consistent with Crohn's disease.
DRGLNOT01	pINCY	Library was constructed using RNA isolated from dorsal root ganglion tissue removed from the cervical spine of a 32-year-old Caucasian male who died from acute pulmonary edema and bronchopneumonia, bilateral pleural and pericardial effusions, and malignant lymphoma (natural killer cell type). Patient history included probable cytomegalovirus, infection, hepatic congestion and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, and Bell's palsy. Surgeries included colonoscopy, large intestine biopsy, adenotonsillectomy, and nasopharyngeal endoscopy and biopsy; treatment included radiation therapy.
MCLDXT02	pINCY	Library was constructed using RNA isolated from treated umbilical cord blood dendritic cells removed from a male. The cells were treated with granulocyte/macrophage colony stimulating factor (GM-CSF), tumor necrosis factor alpha (TNF alpha), stem cell factor (SCF), phorbol myristate acetate (PMA), and ionomycin. The GM-CSF was added at time 0 at 100 ng/ml, the TNF alpha was added at time 0 at 2.5 ng/ml, the SCF was added at time 0 at 25 ng/ml. The PMA and ionomycin were added at 13 days for five hours. Incubation time was 13 days.
MUSLTDR02	PCDNA2.1	This random primed library was constructed using RNA isolated from right lower thigh muscle tissue removed from a 58-year-old Caucasian male during a wide resection of the right posterior thigh. Pathology indicated no residual tumor was identified in the right posterior thigh soft tissue. Changes were consistent with a previous biopsy site. On section through the soft tissue and muscle there was a smooth cystic cavity with hemorrhage around the margin on one side. The wall of the cyst was smooth and pale-tan. Pathology for the matched tumor tissue indicated a grade II liposarcoma. Patient history included liposarcoma (right thigh), and hypercholesterolemia. Previous surgeries included resection of right thigh mass. Family history included myocardial infarction and an unspecified rare blood disease.

Table 6

Library	Vector	Library Description
PITUNOT06	pINCY	Library was constructed using RNA isolated from pituitary gland tissue removed from a 55-year-old male who died from chronic obstructive pulmonary disease. Neuropathology indicated there were no gross abnormalities, other than mild ventricular enlargement. There was no apparent microscopic abnormality in any of the neocortical areas examined, except for a number of silver positive neurons with apical dendrite staining, particularly in the frontal lobe. The significance of this was undetermined. The only other microscopic abnormality was that there was prominent silver staining with some swollen axons in the CA3 region of the anterior and posterior hippocampus. Microscopic sections of the cerebellum revealed mild Bergmann's gliosis in the Purkinje cell layer. Patient history included schizophrenia.
PLACFER06	pINCY	This random primed library was constructed using RNA isolated from placental tissue removed from a Caucasian fetus who died after 16 weeks' gestation from fetal demise and hydrocephalus. Patient history included umbilical cord wrapped around the head (3 times) and the shoulders (1 time). Serology was positive for anti-CMV. Family history included multiple pregnancies and live births, and an abortion.
PROSTUS23	pINCY	This subtracted prostate tumor library was constructed using 10 million clones from a pooled prostate tumor library that was subjected to 2 rounds of subtractive hybridization with 10 million clones from a pooled prostate tissue library. The starting library for subtraction was constructed by pooling equal numbers of clones from 4 prostate tumor libraries using mRNA isolated from prostate tumor removed from Caucasian males at ages 58 (A), 61 (B), 66 (C), and 68 (D) during prostatectomy with lymph node excision. Pathology indicated adenocarcinoma in all donors. History included elevated PSA, induration and tobacco abuse in donor A; elevated PSA, induration, prostate hyperplasia, renal failure, osteoarthritis, renal artery stenosis, benign HTN, thrombocytopenia, hyperlipidemia, tobacco/alcohol abuse and hepatitis C (carrier) in donor B; elevated PSA, induration, and tobacco abuse in donor C; and elevated PSA, induration, hypercholesterolemia, and kidney calculus in donor D. The hybridization probe for subtraction was constructed by pooling equal numbers of cDNA clones from 3 prostate tissue libraries derived from prostate tissue, prostate epithelial cells, and fibroblasts from prostate tissue.
SCORNOT04	pINCY	Library was constructed using RNA isolated from cervical spinal cord tissue removed from a 32-year-old Caucasian male who died from acute pulmonary edema and bronchopneumonia, bilateral pleural and pericardial effusions, and malignant lymphoma (natural killer cell type). Patient history included probable cytomegalovirus, infection, hepatic congestion and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, and Bell's palsy. Surgeries included colonoscopy, large intestine biopsy, adenotonsillectomy, and nasopharyngeal endoscopy and biopsy; treatment included radiation therapy.



Table 6

Library	Vector	Library Description
SINITME01	pINCY	This 5' biased random primed library was constructed using RNA isolated from ileum tissue removed from a 70-year-old Caucasian female during right hemicolectomy, open liver biopsy, flexible sigmoidoscopy, colonoscopy, and permanent colostomy. Pathology for the matched tumor tissue indicated invasive grade 2 adenocarcinoma forming an ulcerated mass, situated 2 cm distal to the ileocecal valve. Patient history included a malignant breast neoplasm, type II diabetes, hyperlipidemia, viral hepatitis, an unspecified thyroid disorder, osteoarthritis, a malignant skin neoplasm, deficiency anemia, and normal delivery. Family history included breast cancer, atherosclerotic coronary artery disease, benign hypertension, cerebrovascular disease, ovarian cancer, and hyperlipidemia.
STOMFET02	pINCY	Library was constructed using RNA isolated from stomach tissue removed from a Hispanic male fetus, who died at 18 weeks' gestation.
TESTNOC01	PBLUESCRIPT	This large size fractionated library was constructed using RNA isolated from testicular tissue removed from a pool of eleven, 10 to 61-year-old Caucasian males.
UTRSTMR02	PCDNA2.1	This random primed library was constructed using pooled cDNA from two different donors. cDNA was generated using mRNA isolated from endometrial tissue removed from a 32-year-old female (donor A) and using mRNA isolated from myometrium removed from a 45-year-old female (donor B) during vaginal hysterectomy and bilateral salpingo-oophorectomy. In donor A, pathology indicated the endometrium was secretory phase. The cervix showed severe dysplasia (CIN III) focally involving the squamocolumnar junction at the 1, 6 and 7 o'clock positions. Mild koilocytotic dysplasia was also identified within the cervix. In donor B, pathology for the matched tumor tissue indicated multiple (23) subserosal, intramural, and submucosal leiomyomata. Patient history included stress incontinence, extrinsic asthma without status asthmaticus and normal delivery in donor B. Family history included cerebrovascular disease, depression, and atherosclerotic coronary artery disease in donor B.

Table 7

Program	Description	Reference	Parameter Threshold
ABI/FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less; Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6; Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less; Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less

Table 7

Program	Description	Reference	Parameter Threshold
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value=1.0E-3 or less; Signal peptide hits: Score= 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score $\geq$ GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater

Table 7

Program	Description	Reference	Parameter Threshold
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
  - a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22,
  - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22,
  - c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and
  - d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.
2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.
3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method of producing a polypeptide of claim 1, the method comprising:
  - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant

polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

- b) recovering the polypeptide so expressed.

5        10. A method of claim 9, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

10        12. An isolated polynucleotide selected from the group consisting of:

- a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44,
- b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ
- 15        ID NO:23-44,
- c) a polynucleotide complementary to a polynucleotide of a),
- d) a polynucleotide complementary to a polynucleotide of b), and
- e) an RNA equivalent of a)-d).

20        13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

- 25        a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
- 30        b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
- 5 b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

10

18. A composition of claim 17, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.

19. A method for treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment the composition of claim 17.

15

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

20

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

25

22. A method for treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

30

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional PKIN, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

- 10 a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- 20 c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

25

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- 30 a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.



29. A method of assessing toxicity of a test compound, the method comprising:
- a) treating a biological sample containing nucleic acids with the test compound,
  - b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,
  - c) quantifying the amount of hybridization complex, and
  - d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.
30. A diagnostic test for a condition or disease associated with the expression of PKIN in a biological sample, the method comprising:
- a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and
  - b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.
31. The antibody of claim 11, wherein the antibody is:
- a) a chimeric antibody,
  - b) a single chain antibody,
  - c) a Fab fragment,
  - d) a F(ab')<sub>2</sub> fragment, or
  - e) a humanized antibody.
32. A composition comprising an antibody of claim 11 and an acceptable excipient.
33. A method of diagnosing a condition or disease associated with the expression of PKIN in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of PKIN in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

5

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibodies from said animal, and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.

10  
15

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

20 39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibody producing cells from the animal,
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
- d) culturing the hybridoma cells, and
- e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.

25  
30

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

5

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22 in a sample, the method comprising:

- 10
- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
  - b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22 in the sample.
- 15

45. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22 from a sample, the method comprising:

- 20
- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
  - b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.

25 46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating a transcript image of a sample which contains polynucleotides, the method comprising:

- 30
- a) labeling the polynucleotides of the sample,
  - b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and
  - c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

5

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is  
10 completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

15 52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

20 54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical  
25 location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

30

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

5 61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

10

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

15 66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

20

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

25 71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

30

74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

76. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.

77. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:22.

5 78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:23.

79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.

80. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:25.

10

81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:26.

82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:27.

15

83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:28.

84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:29.

85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:30.

20

86. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:31.

87. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:32.

25

88. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:33.

89. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:34.

90. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:35.

30

91. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:36.

92. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:37.

93. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:38.

94. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:39.

5 95. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:40.

96. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:41.

97. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:42.

10

98. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:43.

99. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:44.

<110> INCYTE GENOMICS, INC.  
GURURAJAN, Rajagopal  
BAUGHN, Mariah R.  
WALIA, Narinder K.  
ELLIOTT, Vicki S.  
XU, Yuming  
ARVIZU, Chandra  
YAO, Monique G.  
RAMKUMAR, Jayalaxmi  
DING, Li  
TANG, Y. Tom  
HAFALIA, April J.A.  
NGUYEN, Danniel B.  
GANDHI, Ameena R.  
LU, Yan  
YUE, Henry  
BURFORD, Neil  
BANDMAN, Olga  
TRIBOULEY, Catherine  
LAL, Preeti G.  
RECIPON, Shirley A.  
LU, Dyung Aina M.  
BOROWSKY, Mark L.  
THORNTON, Michael  
SWARNAKER Anita  
THANGAVELU, Kavitha  
KHAN, Farrah A.  
ISON, Craig H.

<120> HUMAN KINASES

<130> PI-0262 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/242,410; 60/244,068; 60/245,708; 60/247,672; 60/249,565;  
60/252,730; 60/250,807

<151> 2000-10-20; 2000-10-27; 2000-11-03; 2000-11-09; 2000-11-16  
2000-11-22; 2000-12-01

<160> 44

<170> PERL Program

<210> 1

<211> 337

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7482896CD1

<400> 1

Met Thr Asn Asn Ser Gly Ser Lys Ala Glu Leu Val Val Gly Gly  
1 5 10 15



Ser Ser Pro Ala Pro Pro Glu Gly Pro Ser His Pro Arg Glu Pro  
 1970 1975 1980  
 Ser Thr Pro His Arg Tyr Arg Glu Gly Arg Thr Glu Leu Arg Arg  
 1985 1990 1995  
 Asp Lys Ser Pro Gly Arg Pro Leu Glu Arg Glu Lys Ser Pro Gly  
 2000 2005 2010  
 Arg Met Leu Ser Thr Arg Arg Glu Arg Ser Pro Gly Arg Leu Phe  
 2015 2020 2025  
 Glu Asp Ser Ser Arg Gly Arg Leu Pro Ala Gly Ala Val Arg Thr  
 2030 2035 2040  
 Pro Leu Ser Gln Val Asn Lys Val Trp Asp Gln Ser Ser Val  
 2045 2050

<210> 22

<211> 1665

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7638121CD1

<400> 22

Met Gly Cys Cys Arg Leu Gly Cys Gly Gly Cys Ser Val Ala His  
 1 5 10 15  
 Ser Val Ser Gln Gly Leu Thr Asn His Pro Ser Met Val Gly Cys  
 20 25 30  
 Gly Trp His Pro Gly Leu Cys Gly Trp Gly Gly Gly Leu His Ser  
 35 40 45  
 Ser Leu Pro Ala Leu Pro Gly Pro Pro Ser Met Gln Val Thr Ile  
 50 55 60  
 Glu Asp Val Gln Ala Gln Thr Gly Gly Thr Ala Gln Phe Glu Ala  
 65 70 75  
 Ile Ile Glu Gly Asp Pro Gln Pro Ser Val Thr Trp Tyr Lys Asp  
 80 85 90  
 Ser Val Gln Leu Val Asp Ser Thr Arg Leu Ser Gln Gln Gln Glu  
 95 100 105  
 Gly Thr Thr Tyr Ser Leu Val Leu Arg His Met Ala Ser Lys Asp  
 110 115 120  
 Ala Gly Val Tyr Thr Cys Leu Ala Gln Asn Thr Gly Gly Gln Val  
 125 130 135  
 Leu Cys Lys Ala Glu Leu Leu Val Leu Gly Gly Asp Asn Glu Pro  
 140 145 150  
 Asp Ser Glu Lys Gln Ser His Arg Arg Lys Leu His Ser Phe Tyr  
 155 160 165  
 Glu Val Lys Glu Glu Ile Gly Arg Gly Val Phe Gly Phe Val Lys  
 170 175 180  
 Arg Val Gln His Lys Gly Asn Lys Ile Leu Cys Ala Ala Lys Phe  
 185 190 195  
 Ile Pro Leu Arg Ser Arg Thr Arg Ala Gln Ala Tyr Arg Glu Arg  
 200 205 210  
 Asp Ile Leu Ala Ala Leu Ser His Pro Leu Val Thr Gly Leu Leu  
 215 220 225  
 Asp Gln Phe Glu Thr Arg Lys Thr Leu Ile Leu Ile Leu Glu Leu  
 230 235 240  
 Cys Ser Ser Glu Glu Leu Leu Asp Arg Leu Tyr Arg Lys Gly Val

	245	250	255
Val Thr Glu Ala	Glu Val Lys Val Tyr	Ile Gln Gln Leu Val	Glu
	260	265	270
Gly Leu His Tyr	Leu His Ser His Gly	Val Leu His Leu Asp	Ile
	275	280	285
Lys Pro Ser Asn	Ile Leu Met Val His	Pro Ala Arg Glu Asp	Ile
	290	295	300
Lys Ile Cys Asp	Phe Gly Phe Ala Gln	Asn Ile Thr Pro Ala	Glu
	305	310	315
Leu Gln Phe Ser	Gln Tyr Gly Ser Pro	Glu Phe Val Ser Pro	Glu
	320	325	330
Ile Ile Gln Gln	Asn Pro Val Ser Glu	Ala Ser Asp Ile Trp	Ala
	335	340	345
Met Gly Val Ile	Ser Tyr Leu Ser Leu	Thr Cys Ser Ser Pro	Phe
	350	355	360
Ala Gly Glu Ser	Asp Arg Ala Thr Leu	Leu Asn Val Leu Glu	Gly
	365	370	375
Arg Val Ser Trp	Ser Ser Pro Met Ala	Ala His Leu Ser Glu	Asp
	380	385	390
Ala Lys Asp Phe	Ile Lys Ala Thr Leu	Gln Arg Ala Pro Gln	Ala
	395	400	405
Arg Pro Ser Ala	Ala Gln Cys Leu Ser	His Pro Trp Phe Leu	Lys
	410	415	420
Ser Met Pro Ala	Glu Glu Ala His Phe	Ile Asn Thr Lys Gln	Leu
	425	430	435
Lys Phe Leu Leu	Ala Arg Ser Arg Trp	Gln Arg Ser Leu Met	Ser
	440	445	450
Tyr Lys Ser Ile	Leu Val Met Arg Ser	Ile Pro Glu Leu Leu	Arg
	455	460	465
Gly Pro Pro Asp	Ser Pro Ser Leu Gly	Val Ala Arg His Leu	Cys
	470	475	480
Arg Asp Thr Gly	Gly Ser Ser Ser Ser	Ser Ser Ser Ser Asp	Asn
	485	490	495
Glu Leu Ala Pro	Phe Ala Arg Ala Lys	Ser Leu Pro Pro Ser	Pro
	500	505	510
Val Thr His Ser	Pro Leu Leu His Pro	Arg Gly Phe Leu Arg	Pro
	515	520	525
Ser Ala Ser Leu	Pro Glu Glu Ala Glu	Ala Ser Glu Arg Ser	Thr
	530	535	540
Glu Ala Pro Ala	Pro Pro Ala Ser Pro	Glu Gly Ala Gly Pro	Pro
	545	550	555
Ala Ala Gln Gly	Cys Val Pro Arg His	Ser Val Ile Arg Ser	Leu
	560	565	570
Phe Tyr His Gln	Ala Gly Glu Ser Pro	Glu His Gly Ala Leu	Ala
	575	580	585
Pro Gly Ser Arg	Arg His Pro Ala Arg	Arg Arg His Leu Leu	Lys
	590	595	600
Gly Gly Tyr Ile	Ala Gly Ala Leu Pro	Gly Leu Arg Glu Pro	Leu
	605	610	615
Met Glu His Arg	Val Leu Glu Glu Glu	Ala Ala Arg Glu Glu	Gln
	620	625	630
Ala Thr Leu Leu	Ala Lys Ala Pro Ser	Phe Glu Thr Ala Leu	Arg
	635	640	645
Leu Pro Ala Ser	Gly Thr His Leu Ala	Pro Gly His Ser His	Ser
	650	655	660
Leu Glu His Asp	Ser Pro Ser Thr Pro	Arg Pro Ser Ser Glu	Ala

	665	670	675
Cys Gly Glu Ala	Gln Arg Leu Pro Ser	Ala Pro Ser Gly Gly	Ala
	680	685	690
Pro Ile Arg Asp	Met Gly His Pro Gln	Gly Ser Lys Gln Leu	Pro
	695	700	705
Ser Thr Gly Gly	His Pro Gly Thr Ala	Gln Pro Glu Arg Pro	Ser
	710	715	720
Pro Asp Ser Pro	Trp Gly Gln Pro Ala	Pro Phe Cys His Pro	Lys
	725	730	735
Gln Gly Ser Ala	Pro Gln Glu Gly Cys	Ser Pro His Pro Ala	Val
	740	745	750
Ala Pro Cys Pro	Pro Gly Ser Phe Pro	Pro Gly Ser Cys Lys	Glu
	755	760	765
Ala Pro Leu Val	Pro Ser Ser Pro Phe	Leu Gly Gln Pro Gln	Ala
	770	775	780
Pro Leu Ala Pro	Ala Lys Ala Ser Pro	Pro Leu Asp Ser Lys	Met
	785	790	795
Gly Pro Gly Asp	Ile Ser Leu Pro Gly	Arg Pro Lys Pro Gly	Pro
	800	805	810
Cys Ser Ser Pro	Gly Ser Ala Ser Gln	Ala Ser Ser Ser Gln	Val
	815	820	825
Ser Ser Leu Arg	Val Gly Ser Ser Gln	Val Gly Thr Glu Pro	Gly
	830	835	840
Pro Ser Leu Asp	Ala Glu Gly Trp Thr	Gln Glu Ala Glu Asp	Leu
	845	850	855
Ser Asp Ser Thr	Pro Thr Leu Gln Arg	Pro Gln Glu Gln Val	Thr
	860	865	870
Met Arg Lys Phe	Ser Leu Gly Gly Arg	Gly Gly Tyr Ala Gly	Val
	875	880	885
Ala Gly Tyr Gly	Thr Phe Ala Phe Gly	Gly Asp Ala Gly Gly	Met
	890	895	900
Leu Gly Gln Gly	Pro Met Trp Ala Arg	Ile Ala Trp Ala Val	Ser
	905	910	915
Gln Ser Glu Glu	Glu Glu Gln Glu Glu	Ala Arg Ala Glu Ser	Gln
	920	925	930
Ser Glu Glu Gln	Gln Glu Ala Arg Ala	Glu Ser Pro Leu Pro	Gln
	935	940	945
Val Ser Ala Arg	Pro Val Pro Glu Val	Gly Arg Ala Pro Thr	Arg
	950	955	960
Ser Ser Pro Glu	Pro Thr Pro Trp Glu	Asp Ile Gly Gln Val	Ser
	965	970	975
Leu Val Gln Ile	Arg Asp Leu Ser Gly	Asp Ala Glu Ala Ala	Asp
	980	985	990
Thr Ile Ser Leu	Asp Ile Ser Glu Val	Asp Pro Ala Tyr Leu	Asn
	995	1000	1005
Leu Ser Asp Leu	Tyr Asp Ile Lys Tyr	Leu Pro Phe Glu Phe	Met
	1010	1015	1020
Ile Phe Arg Lys	Val Pro Lys Ser Ala	Gln Pro Glu Pro Pro	Ser
	1025	1030	1035
Pro Met Ala Glu	Glu Glu Leu Ala Glu	Phe Pro Glu Pro Thr	Trp
	1040	1045	1050
Pro Trp Pro Gly	Glu Leu Gly Pro His	Ala Gly Leu Glu Ile	Thr
	1055	1060	1065
Glu Glu Ser Glu	Asp Val Asp Ala Leu	Leu Ala Glu Ala Ala	Val
	1070	1075	1080
Gly Arg Lys Arg	Lys Trp Ser Ser Pro	Ser Arg Ser Leu Phe	His

1085	1090	1095
Phe Pro Gly Arg His Leu Pro Leu Asp Glu Pro Ala Glu Leu Gly		
1100	1105	1110
Leu Arg Glu Arg Val Lys Ala Ser Val Glu His Ile Ser Arg Ile		
1115	1120	1125
Leu Lys Gly Arg Pro Glu Gly Leu Glu Lys Glu Gly Pro Pro Arg		
1130	1135	1140
Lys Lys Pro Gly Leu Ala Ser Phe Arg Leu Ser Gly Leu Lys Ser		
1145	1150	1155
Trp Asp Arg Ala Pro Thr Phe Leu Arg Glu Leu Ser Asp Glu Thr		
1160	1165	1170
Val Val Leu Gly Gln Ser Val Thr Leu Ala Cys Gln Val Ser Ala		
1175	1180	1185
Gln Pro Ala Ala Gln Ala Thr Trp Ser Lys Asp Gly Ala Pro Leu		
1190	1195	1200
Glu Ser Ser Ser Arg Val Leu Ile Ser Ala Thr Leu Lys Asn Phe		
1205	1210	1215
Gln Leu Leu Thr Ile Leu Val Val Val Ala Glu Asp Leu Gly Val		
1220	1225	1230
Tyr Thr Cys Ser Val Ser Asn Ala Leu Gly Thr Val Thr Thr Thr		
1235	1240	1245
Gly Val Leu Arg Lys Ala Glu Arg Pro Ser Ser Ser Pro Cys Pro		
1250	1255	1260
Asp Ile Gly Glu Val Tyr Ala Asp Gly Val Leu Leu Val Trp Lys		
1265	1270	1275
Pro Val Glu Ser Tyr Gly Pro Val Thr Tyr Ile Val Gln Cys Ser		
1280	1285	1290
Leu Glu Gly Gly Ser Trp Thr Thr Leu Ala Ser Asp Ile Phe Asp		
1295	1300	1305
Cys Cys Tyr Leu Thr Ser Lys Leu Ser Arg Gly Gly Thr Tyr Thr		
1310	1315	1320
Phe Arg Thr Ala Cys Val Ser Lys Ala Gly Met Gly Pro Tyr Ser		
1325	1330	1335
Ser Pro Ser Glu Gln Val Leu Leu Gly Gly Pro Ser His Leu Ala		
1340	1345	1350
Ser Glu Glu Glu Ser Gln Gly Arg Ser Ala Gln Pro Leu Pro Ser		
1355	1360	1365
Thr Lys Thr Phe Ala Phe Gln Thr Gln Ile Gln Arg Gly Arg Phe		
1370	1375	1380
Ser Val Val Arg Gln Cys Trp Glu Lys Ala Ser Gly Arg Ala Leu		
1385	1390	1395
Ala Ala Lys Ile Ile Pro Tyr His Pro Lys Asp Lys Thr Ala Val		
1400	1405	1410
Leu Arg Glu Tyr Glu Ala Leu Lys Gly Leu Arg His Pro His Leu		
1415	1420	1425
Ala Gln Leu His Ala Ala Tyr Leu Ser Pro Arg His Leu Val Leu		
1430	1435	1440
Ile Leu Glu Leu Cys Ser Gly Pro Glu Leu Leu Pro Cys Leu Ala		
1445	1450	1455
Glu Arg Ala Ser Tyr Ser Glu Ser Glu Val Lys Asp Tyr Leu Trp		
1460	1465	1470
Gln Met Leu Ser Ala Thr Gln Tyr Leu His Asn Gln His Ile Leu		
1475	1480	1485
His Leu Asp Leu Arg Ser Glu Asn Met Ile Ile Thr Glu Tyr Asn		
1490	1495	1500
Leu Leu Lys Val Val Asp Leu Gly Asn Ala Gln Ser Leu Ser Gln		

1505	1510	1515
Glu Lys Val Leu Pro Ser Asp Lys Phe Lys Asp Tyr Leu Glu Thr		
1520	1525	1530
Met Ala Pro Glu Leu Leu Glu Gly Gln Gly Ala Val Pro Gln Thr		
1535	1540	1545
Asp Ile Trp Ala Ile Gly Val Thr Ala Phe Ile Met Leu Ser Ala		
1550	1555	1560
Glu Tyr Pro Val Ser Ser Glu Gly Ala Arg Asp Leu Gln Arg Gly		
1565	1570	1575
Leu Arg Lys Gly Leu Val Arg Leu Ser Arg Cys Tyr Ala Gly Leu		
1580	1585	1590
Ser Gly Gly Ala Val Ala Phe Leu Arg Ser Thr Leu Cys Ala Gln		
1595	1600	1605
Pro Trp Gly Arg Pro Cys Ala Ser Ser Cys Leu Gln Cys Pro Trp		
1610	1615	1620
Leu Thr Glu Glu Gly Pro Ala Cys Ser Arg Pro Ala Pro Val Thr		
1625	1630	1635
Phe Pro Thr Ala Arg Leu Arg Val Phe Val Arg Asn Arg Glu Lys		
1640	1645	1650
Arg Arg Ala Leu Leu Tyr Lys Arg His Asn Leu Ala Gln Val Arg		
1655	1660	1665

&lt;210&gt; 23

&lt;211&gt; 1014

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7482896CB1

&lt;400&gt; 23

```

atgacaaaaca acagcgggtc caaagccgaa ctggttgtgg gagggaaata caaactgggtg 60
cggaagatcg ggtctggctc ctttggagac gtttatctgg gcatcaccac caccaacggc 120
gaggacgtag cagtgaagct ggaatctcag aagggtcaagc acccccagtt gctgtatgag 180
agcaaaactct acacgattct tcaagggtgg gttggcatcc cccacatgca ctggtatggt 240
caggaaaaaag acaacaatgt gctagtcatt gaccttctgg gacccagcct cgaagacctc 300
tttaatttct gttcaagaag gttcaccatg aaaactgtac ttatgtagc cgaccagatg 360
atcagcagaa ttgaatacgt gcatacaaag aattttctac accgagacat taaaccagat 420
aacttcctga tgggtactgg gcgtcactgt aataagttgt tccttattga ttttggtttg 480
gccaaaaagt acagagacaa caggaccagg caacacatac cgtacagaga agataaacac 540
ctcattggca ctgtccgata tgccagcatc aatgcacatc ttggtattga gcagagccgc 600
cgagatgaca tggaatcctt aggtacgtt ttcattgatt ttaatagaac cagcctgccg 660
tggaaggac taagggtat gacaaaaaaa caaaaatatg aaaagattag tgagaagaag 720
atgtccaccc ctgttgaagt tttatgtaag gggtttcctg cagaattcgc catgtacttg 780
aactactgtc gtgggctgct ctttgaggaa gtcccagatt acatgtatct gaggcagcta 840
ttccgcattc ttttcaggac cctgaaccac caatatgact acacatttga ttggacgatg 900
ttaaagcaga aagcagcaca gcaggcagcc tcttcagtg ggcagggtca gcaggcccaa 960
accagacag gcaagcaaac tgaaaaaac aagaataatg tgaaagataa ctaa 1014

```

&lt;210&gt; 24

&lt;211&gt; 1530

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

```

ccccaccgct accgcgaggg ggcgaccgag ctgcgcaggg acaagtctcc tggccgcccc 6060
ctggagcgag agaagtcccc cggccggatg ctcagcacgc ggagagagcg gtcccccggg 6120
aggctgtttg aagacagcag caggggcccg ctgcctgcgg gagccgtgag gacccccgctg 6180
tcccagggtga acaaggtctg ggaccagtct tcagtataaa tctcagccag aaaaaccaac 6240
tcctcatctt gatctgcagg aaaacaccaa acacactatg gaactctgct gatggggga 6298

```

&lt;210&gt; 44

&lt;211&gt; 5454

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7638121CB1

&lt;400&gt; 44

```

cacgcacacc gcacgtacgg ggttggggccc agctgggtta taagcgtgat ccccatgccc 60
cctgcccagg ctgggggggca tttgcacatc tgcaaaggcc tcccagcctg tcccagccct 120
gccccagcct gggacccccca cattctactc accgtgtctc ctcagagggg ccagaaccct 180
ccactgggga gaggcaagtg gcggtgaact tgggtgtccat aggaccctgt ccctgagagc 240
gacagctgag ttagtgagct ccactggccc caccaactcc ttctgatcac ctggccagct 300
gaggtcagag tgggagaggc agtgggtcca ttgaaggagt actcctaact gtcagaagcc 360
tgggcgggtca ggatgggggtg ctgtcgcttg ggctgcgggg ggtgttcagt tgcccacagt 420
gtatctcagg gtctcaccaa ccatccaagc atggtaggct gtggctggca cccaggggtg 480
tgtggctggg gaggtgggtc ccacagttcc ctccctgccc tcccagggcc cccatccatg 540
caggtaacca tcgaggatgt gcaggcacag acaggcggaa cggcccaatt cgaggctatc 600
attgagggcg acccacagcc ctcggtgacc tggtaacaagg acagcgtcca gctggtggac 660
agcaccgggc ttagccagca gcaagaaggc accacatact ccctggtgct gaggcatatg 720
gcctcgaagg atgccggcgt ttacacctgc ctggcccaaa aactggtgg ccagggtgctc 780
tgcaaggcag agctgctggt gcttgggggg gacaatgagc cggactcaga gaagcaaagc 840
caccggagga agctgcactc cttctatgag gtcaaggagg agattggaag gggcgtgttt 900
ggcttcgtaa aaagagtgca gcacaaagga aacaagatct tgtgcgctgc caagttcatc 960
cccctacgga gcagaactcg ggcccaggca tacagggagc gagacatcct ggccgcgctg 1020
agccaccgca tgggtcacggg gctgctggac cagtttgaga cccgcaagac cctcatcctc 1080
atcctggagc tgtgtctcatc cgaggagctg ctggaccgcc tgtacaggaa gggcgtggtg 1140
acggaggccg aggtcaaggt ctacatccag cagctggtgg aggggctgca ctacctgcac 1200
agccatggcg ttctccacct ggacataaag cctctaaaca tcctgatggt gcatcctgcc 1260
cgggaagaca ttaaaatctg cgactttggc tttgcccaga acatcacccc agcagagctg 1320
cagttcagcc agtacggctc ccctgagttc gtctcccccg agatcateca gcagaaccct 1380
gtgagcgaag cctccgacat ttggggccatg ggtgtcatct cctacctcag cctgacctgc 1440
tcatcccat ttgccggcga gagtgaaccg gccaccctcc tgaacgtcct ggaggggccc 1500
gtgtcatgga gcagcccat ggctgcccac ctcagcgaag acgcccaga cttcatcaag 1560
gctacgctgc agagagcccc tcaggcccgg cctagtgcgg cccagtgcct ctcccacccc 1620
tggttcctga aatccatgcc tgccggaggag gcccaactca tcaacaccaa gcagctcaag 1680
ttctcctggt ccgaagtcg ctggcagcgt tccctgatga gctacaagtc catcctggtg 1740
atgcgctcca tccctgagct gctgcggggc ccacccgaca gccctcctc cggcgtagcc 1800
cggcacctct cgagggacac tgggtggtcc tccagtccct cctcctcctc tgacaacgag 1860
ctcgccccat ttgcccgggc taagtactg ccaccctccc cggtgacaca ctcaccatg 1920
ctgaccccc ggggttctct ggggccctcg gccagcctgc ctgaggaagc cgaggccagt 1980
gagcgctcca ccgaggcccc agctccgcct gcatctcccg agggtgccgg gccaccggcc 2040
gccaggggct gcgtgccccg gcacagcgtc atccgcagcc tgttctacca ccaggcgggt 2100
gagagccctg agcacggggc cctggccccg gggagcaggg gccaccgggc ccggcggcgg 2160
cacctgctga agggcgggta cattgcgggg gcgctgccag gcctgcgcga gccactgatg 2220
gagcacggcg tgctggagga ggaggccgcc agggaggagc agggccaccct cctggccaaa 2280
gccccctcat tcgagactgc cctcggctg cctgcctctg gcacccactt ggccccctggc 2340
cacagccact ccctggaaca tgactctccg agcacccccc gccctcctc ggaggccctgc 2400

```

```

ggtagggcac agcgactgcc ttcagcccc tccggggggg cccctatcag ggacatgggg 2460
caccctcagg gctccaagca gcttccatcc actgggtggcc acccaggcac tgctcagcca 2520
gagaggccat ccccgagcag cccttggggg cagccagccc ctttctgccca ccccaagcag 2580
ggttctgccc cccaggaggg ctgcagcccc caccagcag ttgccccatg ccctcctggc 2640
tccttccctc caggatcttg caaagaggcc cccttagtac cctcaagccc cttcttggga 2700
cagccccagg cacccttgc ccctgcaaaa gcaagcccc cattggactc taagatgggg 2760
cctggagaca tctctcttcc tgggaggcca aaaccggcc cctgcagtcc cccagggtca 2820
gctcccagg cgagctcttc ccaagtgagc tccctcaggg tgggctcctc ccagggtggc 2880
acagagcctg gcccctccct ggatgaggag ggctggaccc aggaggctga ggatctgtcc 2940
gactccacac ccaccttga gggcctcag gaacagggtga ccatgcgcaa gttctccctg 3000
ggtggtcgcg ggggtacgc aggcgtggct ggctatggca cctttgcctt tgggtggagat 3060
gcagggggca tgctggggca ggggcccatt tgggcccagga tagcctgggc tgtgtccag 3120
tcggaggagg aggagcagga ggaggccagg gctgagtcct agtcggagga gcagcaggag 3180
gccagggctg agagcccact gcccagggtc agtgcaaggc ctgtgcctga ggtcggcagg 3240
gtccccacca ggagctctcc agagcccacc ccatgggagg acatcgggca ggtctccctg 3300
gtgcagatcc gggacctgtc aggtgatgag gaggcggccg acacaatatc cctggacatt 3360
tccgaggtgg accccgccta cctcaacctc tcagacctgt acgatatcaa gtacctcca 3420
ttcgagttta tgatcttcag gaaagtcccc aagtcgcgtc agccagagcc gccctcccc 3480
atggctgagg aggagctggc cgagttcccg gaccccaggt ggccctggcc aggtgaactg 3540
ggccccacg caggcctgga gatcacagag gagtcaagg atgtggacgc gctgctggca 3600
gaggctgccg tgggcaggaa gcgcaagtgg tcctcgccgt cagcagcct cttccacttc 3660
cctgggaggc acctgccgt ggacgagcct gcagagctgg ggctgcgtga gagagtgaag 3720
gcctccgtgg agcacatctc ccggtacctg aagggcaggc cggaaggtct ggagaaggag 3780
gggcccccca ggaagaagcc aggccttgct tccttccggc tctcaggtct gaagagctgg 3840
gaccgagcgc cgacattcct aagggagctc tcagatgaga ctgtgtctct gggccagtca 3900
gtgacactgg cctgccagggt gtcagcccag ccagctgccc aggccacctg gagcaaagac 3960
ggagcccccc tggagagcag cagccgtgtc ctcatctctg ccacctcaa gaacttccag 4020
cttctgacca tcctggtggt ggtggctgag gacctgggtg tgtacacctg cagcgtgagc 4080
aatgcgctgg ggacagtgac caccacgggc gtcctccgga aggcagagcg cccctcatct 4140
tcgccatgcc cggatatcgg ggaggtgtac gcggtgggg tgctgctggt ctggaagccc 4200
gtggaatcct acggccctgt gacctacatt gtgcagtga gcctagaagg cggcagctgg 4260
accacactgg cctccgacat ctttgactgc tgctacctga ccagcaagct ctcccggggt 4320
ggcacctaca cttccgcac ggcattgtgc agcaaggcag gaatgggtcc ctacagcagc 4380
ccctcggagc aagtcctcct gggaggggccc agccacctgg cctctgagga ggagagccag 4440
gggcggtcag cccaaccct gcccagcaca aagacctcg cattccagac acagatccag 4500
aggggcccgt tcagcgtggt gcggcaatgc tgggagaagg ccagcggggc ggcgtggcc 4560
gccaagatca tcccctacca cccaaggac aagacagcag tgctgcgca atacgaggcc 4620
ctcaagggcc tgcgccacc gcacctggcc cagctgcagc cagcctacct cagccccgg 4680
cacctggtgc tcatcttga gctgtgctct gggcccagc tgctcccctg cctggccgag 4740
agggcctcct actcagaatc cgaggtgaag gactacctgt ggcagatgtt gagtgccacc 4800
cagtacctgc acaaccagca catcctgcac ctggacctga ggtccgagaa catgatcatc 4860
accgaataca acctgctcaa ggtcgtggac ctgggcaatg cacagagcct cagccaggag 4920
aaggtgctgc cctcagacaa gttcaaggac tacctagaga ccatggctcc agagctcctg 4980
gagggccagg gggctgttcc acagacagac atctgggcca tcggtgtgac agccttcac 5040
atgctgagcg ccgagtacc ggtgagcagc gagggtgcac gcgacctga gagaggactg 5100
cgcaaggggc tggctccggct gagccgctgc tacgcggggc tgtccggggg cgccgtggcc 5160
ttctgcgca gactctgtg cgcacagccc tggggccggc cctgcgcgtc cagctgctg 5220
cagtgcccgt ggctaacaga ggaggggccc gctgttccg ggcgcgcgc cgtgaccttc 5280
cctaccgcg ggctgcgcgt ctctgtgcg aatcgcgaga agagacgcgc gctgctgtac 5340
aagaggcaca acctggccca ggtgcgctga gggtcgcccc ggccacacc ttggtctccc 5400
cgctgggggt cgctgcagac gcgccaataa aaacgcacag ccgggcgaga aaaa 5454

```

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
25 April 2002 (25.04.2002)

PCT

(10) International Publication Number  
WO 02/033099 A3

(51) International Patent Classification<sup>7</sup>: C12N 15/54,  
9/12, C12Q 1/68, 1/48, A01K 67/027, C07K 16/40, G01N  
33/50, A61K 38/45

(21) International Application Number: PCT/US01/47728

(22) International Filing Date: 20 October 2001 (20.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/242,410 20 October 2000 (20.10.2000) US  
60/244,068 27 October 2000 (27.10.2000) US  
60/245,708 3 November 2000 (03.11.2000) US  
60/247,672 9 November 2000 (09.11.2000) US  
60/249,565 16 November 2000 (16.11.2000) US  
60/252,730 22 November 2000 (22.11.2000) US  
60/250,807 1 December 2000 (01.12.2000) US

(71) Applicant (for all designated States except US): INCYTE  
GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo  
Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): GURURAJAN,  
Rajagopal [US/US]; 5591 Dent Avenue, San Jose, CA  
95118 (US). BAUGHN, Mariah, R. [US/US]; 14244  
Santiago Road, San Leandro, CA 94577 (US). CHAWLA,  
Narinder, K. [US/US]; 33 Union Square, #712, Union  
City, CA 94587 (US). ELLIOTT, Vicki, S. [US/US]; 3770  
Polton Place Way, San Jose, CA 95121 (US). XU, Yuming  
[US/US]; 1739 Walnut Drive, Mountain View, CA 94040  
(US). ARVIZU, Chandra [US/US]; 490 Sherwood Way  
#1, Menlo Park, CA 94025 (US). YAO, Monique, G.  
[US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US).  
RAMKUMAR, Jaya [IN/US]; 34359 Maybird Circle,  
Fremont, CA 94555 (US). DING, Li [CN/US]; 3353 Alma  
Street #146, Palo Alto, CA 94306 (US). TANG, Y., Tom  
[US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US).  
HAFALIA, April, J., A. [US/US]; 2227 Calle de Primavera,  
Santa Clara, CA 95054 (US). NGUYEN, Dannel,  
B. [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118  
(US). GANDHI, Ameena, R. [US/US]; 705 5th Avenue,  
San Francisco, CA 94118 (US). LU, Yan [CN/US];  
3885 Cornnua Way, Palo Alto, CA 94303 (US). YUE,  
Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087  
(US). BURFORD, Neil [GB/US]; 105 Wildwood Circle,

Durham, CT 06422 (US). BANDMAN, Olga [US/US];  
366 Anna Avenue, Mountain View, CA 94043 (US). TRI-  
BOULEY, Catherine, M. [FR/US]; 1121 Tennessee, #5,  
San Francisco, CA 94107 (US). LAL, Preeti, G. [IN/US];  
P.O. Box 5142, Santa Clara, CA 95056 (US). RECIPON,  
Shirley, A. [US/US]; 85 Fortuna Avenue, San Francisco,  
CA 94115 (US). LU, Dyung, Aina, M. [US/US]; 233 Coy  
Drive, San Jose, CA 95123 (US). BOROWSKY, Mark,  
L. [US/US]; 122 Orchard Avenue, Redwood City, CA  
94061 (US). THORNTON, Michael [US/US]; 9 Medway  
Road, Woodside, CA 94062 (US). SWARNAKER, Anita  
[CA/US]; 8 Locksley Avenue, #5D, San Francisco, CA  
94122 (US). THANGAVELU, Kavitha [IN/US]; 1950  
Montecito Avenue #23, Mountain View, CA 94304 (US).  
KHAN, Farrah, A. [IN/US]; 3617 Central Road, #102,  
Glenview, IL 60025 (US). ISON, Craig, H. [US/US];  
1242 Weathersfield Way, San Jose, CA 95118 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics,  
Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,  
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,  
ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,  
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,  
TG).

Published:

- with international search report
- before the expiration of the time limit for amending the  
claims and to be republished in the event of receipt of  
amendments

(88) Date of publication of the international search report:  
9 October 2003

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: HUMAN KINASES

(57) Abstract: The invention provides human human kinases (PKIN) and polynucleotides which identify and encode PKIN. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of PKIN.



WO 02/033099 A3



## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 01/47728

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/54 C12N9/12 C12Q1/68 C12Q1/48 A01K67/027  
 C07K16/40 G01N33/50 A61K38/45

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C12Q A01K C07K G01N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, MEDLINE, SEQUENCE SEARCH, WPI Data, PAJ, EMBL

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	<p>WO 01 38503 A (PLOWMAN GREGORY D ;CLARY DOUGLAS (US); SUGEN INC (US); WHYTE DAVID) 31 May 2001 (2001-05-31)</p> <p>* Sequence "SGK003_ID#NA_22" has 100% identity in a 1014nt overlap to SEQ ID NO.23. Sequence "SGK003_ID#AA_79" has 100% identity to SEQ ID NO.1 over a 337aa overlap *</p> <p>page 125; figures 1,2'</p> <p>* Sequence "SGK058_ID#AA_106" has 100% identity over a 278aa overlap to SEQ ID NO.6. Sequence "SGK058_ID#NA_49" has 100% identity in a 834nt overlap to SEQ ID NO.28 *</p> <p>page 135</p> <p>page 146</p> <p>page 232</p> <p>page 250</p> <p>page 273</p> <p style="text-align: center;">-/-</p>	<p>1-20,23, 26-56, 61,78,83</p>

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

## \* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\*&\* document member of the same patent family

Date of the actual completion of the international search

8 July 2003

Date of mailing of the international search report

04.08.03

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+31-70) 340-2040, Tx. 31 651 epo.nl.  
 Fax: (+31-70) 340-3016

Authorized officer

Meacock, S

## INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 01/47728

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E	<p>page 291, line 14 -page 292, line 8 page 306, line 15 - line 28 ---</p> <p>WO 01 98342 A (GLAXO GROUP LTD ;MURDOCK PAUL R (GB); SMITHKLINE BEECHAM PLC (GB);) 27 December 2001 (2001-12-27) SEQ ID NO.10 shows 100% identity over 1014nt overlap to SEQ ID NO.23 of the present application. SEQ ID NO.32 shows 100% identity over 337aa overlap with SEQ ID NO.1 * page 25 -page 26 page 29 page 47 -page 50 ---</p>	1-20,23, 26-56,78
X	<p>FISH KIMBERLY J ET AL: "Isolation and Characterization of Human Casein Kinase I-epsilon (CKI), a Novel Member of the CKI Gene Family." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 270, no. 25, 1995, pages 14875-14883, XP002177865 ISSN: 0021-9258 cited in the application page 14877, left-hand column; figures 1,2 ---</p>	1-20,23, 26-56,78
X	<p>WO 94 17189 A (SALK INST FOR BIOLOGICAL STUDI) 4 August 1994 (1994-08-04) * SEQ ID NO.11 shows 91% identity in a 1014nt overlap to SEQ ID NO.23 of the present application. SEQ ID NO.23 shows 91% identity over 335aa overlap to SEQ ID NO.1 * page 83 -page 87 ---</p>	1-20,23, 26-56,78
X	<p>ROWLES J ET AL: "PURIFICATION OF CASEIN KINASE I AND ISOLATION OF CDNAS ENCODING MULTIPLE CASEIN KINASE I-LIKE ENZYMES" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES, vol. 88, no. 21, 1991, pages 9548-9552, XP001119756 ISSN: 0027-8424 ---</p>	1,3,12
Y	<p>* Figure 2 sequence shows 91% identity to SEQ ID NO.1 over 324aa overlap, and SEQ ID NO.23 over 971nt overlap * abstract; figure 2 ---</p>	13-20, 23, 26-56,78
	<p>--- -/--</p>	

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 01/47728

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PULGAR VICTOR ET AL: "The recombinant alpha isoform of protein kinase CK1 from <i>Xenopus laevis</i> can phosphorylate tyrosine in synthetic substrates." EUROPEAN JOURNAL OF BIOCHEMISTRY, vol. 242, no. 3, 1996, pages 519-528, XP001119751 ISSN: 0014-2956	1,3,12
Y	* Figure 1 sequence shows 91% identity to SEQ ID NO.1 over 335aa overlap * abstract; figure 1	13-20, 23, 26-56,78
X	GREEN C L ET AL: "Identification of four alternatively spliced isoforms of chicken casein kinase I alpha that are all expressed in diverse cell types" GENE, ELSEVIER BIOMEDICAL PRESS, AMSTERDAM, NL, vol. 216, no. 1, August 1998 (1998-08), pages 189-195, XP004149296 ISSN: 0378-1119	1,3,12
Y	abstract; figure 1 * CKIalphaS sequence shows 91% identity to SEQ ID NO.1 over 335aa overlap * page 193, right-hand column	13-20, 23, 26-56,78
P,X	DATABASE EMBL 'Online! 1 March 2001 (2001-03-01) "Hypothetical protein FLJ23074" Database accession no. Q9H5T2 XP002246601 * 99.4% identity over 168aa overlap to SEQ ID NO.6 * the whole document	1-20,23, 26-55, 61,83
E	WO 02 14355 A (MERCK PATENT GMBH ;SCHARM BURKHARD (DE)) 21 February 2002 (2002-02-21) * SEQ ID NO.1 has 99.8% identity over 1137nt overlap to SEQ ID NO.28 of the present application; SEQ ID NO.4 has 100% identity over 233aa to SEQ ID NO.6 of the present application * abstract; claim 2 page 1 -page 3	1-20,23, 26-55, 61,83

-/--

## INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 01/47728

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GRAVES D J: "Powerful tools for genetic analysis come of age" TRENDS IN BIOTECHNOLOGY, ELSEVIER, AMSTERDAM, NL, vol. 17, no. 3, March 1999 (1999-03), pages 127-134, XP004157733 ISSN: 0167-7799 page 127, right-hand column page 130, left-hand column page 131, right-hand column page 132, left-hand column, paragraph 2	46-55
A	CHUNG CHANG Y ET AL: "A novel, putative MEK kinase controls developmental timing and spatial patterning in Dictyostelium and is regulated by ubiquitin-mediated protein degradation." GENES & DEVELOPMENT, vol. 12, no. 22, 15 November 1998 (1998-11-15), pages 3564-3578, XP002246600 ISSN: 0890-9369 * MEC kinase alpha has 47% identity over 247aa overlap to SEQ ID NO.6 * abstract; figure 1	1-20, 23, 26-55, 61, 83

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US 01/47728

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
see FURTHER INFORMATION sheet PCT/ISA/210
2. ☒ Claims Nos.: 21, 22, 24 and 25  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:  
see FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:  
Inventions 1 and 6 (corresponding to SEQ ID NOs. 1, 6, 23 and 28): Claims 1-55 (all partially) and Claims 56, 61, 78 and 83 (all completely)
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☒ No protest accompanied the payment of additional search fees.

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

## Continuation of Box I.1

Although claim 19 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Although claims 33 and 35 are directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

## Continuation of Box I.2

Claims Nos.: 21, 22, 24 and 25

Present Claims 21, 22, 24 and 25 relate to a compound defined by reference to a desirable characteristic or property, namely ant/agonist activity to the polypeptide of Claim 1.

The cited claims cover all compounds having this characteristic or property, whereas the application provides no support within the meaning of Art. 6 PCT and/or disclosure within the meaning of Art. 5 PCT for such compounds. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible.

Independent of the above reasoning, the claims also lack clarity (Art. 6 PCT). An attempt is made to define the product by reference to a result to be achieved. Again, this lack of clarity in the present case is such as to render a meaningful search over the whole of the claimed scope impossible.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

Invention 1: Claims 1-55 (all partially),  
56 and 78 (both completely)

An isolated polypeptide which comprises the amino acid sequence of SEQ ID NO.1, comprises a polypeptide comprising a naturally-occurring amino acid sequence at least 90% identical to said polypeptide, or is a biologically active or immunogenic fragment of the amino acid sequence of SEQ ID NO.1; a method of producing, purifying or detecting said polypeptide; an isolated antibody which specifically binds said polypeptide; compositions comprising, and methods of making, said antibody; an isolated polynucleotide encoding said polypeptide or fragment; a recombinant polynucleotide comprising a promoter sequence operably linked to said polynucleotide; a cell transformed with, or a transgenic organism comprising, said recombinant polynucleotide; an isolated polynucleotide comprising a naturally occurring sequence at least 90% identical to SEQ ID NO.23; a polynucleotide complementary to said isolated polynucleotide or an RNA equivalent thereof; an isolated polynucleotide comprising at least 60 contiguous nucleotides of said polynucleotide; an array comprising said polynucleotide; methods of assessing toxicity of a test compound using said polynucleotide; a method of screening a compound for effectiveness in altering expression of said polynucleotide; a method of diagnosing a condition or disease using said polynucleotide or antibody; a method of generating a transcript image of a sample; a method of screening a compound for effectiveness as an agonist or antagonist of, or which specifically binds to, or that modulates the activity of, said polypeptide; a composition comprising an agonist or antagonist identified by said screening method and a method of treating a disease using said agonist or antagonist.

Inventions 2-22: Claims 1-55 (all partially),  
Claims 57-77 and 79-99 (all completely)

Idem as for invention 1, but relating to an isolated polypeptide comprising the amino acid sequence of SEQ ID NOs.2-22 with the corresponding nucleotide sequences of SEQ ID NOs.24-44, respectively. Hence, invention 2 relates to SEQ ID NOs.2 and 24, invention 3 relates to SEQ ID NOs.3 and 25, and so on, with invention 22 relating to SEQ ID NOs.22 and 44.

In the interests of conciseness, the subject matter (which includes the associated aspects of encoding polynucleotides, antibodies, methods etc) of the inventions has been defined

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

explicitly only for the first invention, and the additional inventions defined only by analogy thereto.



# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 01/47728

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 0138503	A	31-05-2001	AU 1926001 A	04-06-2001
			CA 2394803 A1	31-05-2001
			EP 1240194 A2	18-09-2002
			JP 2003514583 T	22-04-2003
			WO 0138503 A2	31-05-2001
WO 0198342	A	27-12-2001	AU 6867101 A	02-01-2002
			BR 0111877 A	24-06-2003
			CA 2414005 A1	27-12-2001
			CZ 20024148 A3	14-05-2003
			EP 1292617 A1	19-03-2003
			NO 20026162 A	21-02-2003
			WO 0198342 A1	27-12-2001
WO 9417189	A	04-08-1994	CA 2132452 A1	04-08-1994
			EP 0632832 A1	11-01-1995
			JP 7505057 T	08-06-1995
			WO 9417189 A2	04-08-1994
			US 6060296 A	09-05-2000
			US 5627064 A	06-05-1997
			US 5756289 A	26-05-1998
			US 5686412 A	11-11-1997
WO 0214355	A	21-02-2002	AU 9174501 A	25-02-2002
			WO 0214355 A2	21-02-2002
			EP 1307566 A2	07-05-2003